

Uniform ε -Stability of Distributed Nonlinear Filtering over DNAs: Gaussian-Finite HMMs

Dionysios S. Kalogerias and Athina P. Petropulu

September 2016

Abstract

In this work, we study stability of distributed filtering of Markov chains with finite state space, partially observed in conditionally Gaussian noise. We consider a nonlinear filtering scheme over a Distributed Network of Agents (DNA), which relies on the distributed evaluation of the likelihood part of the centralized nonlinear filter and is based on a particular specialization of the Alternating Direction Method of Multipliers (ADMM) for fast average consensus. Assuming the same number of consensus steps between any two consecutive noisy measurements for each sensor in the network, we fully characterize a minimal number of such steps, such that the distributed filter remains uniformly stable with a prescribed accuracy level, $\varepsilon \in (0, 1]$, within a finite operational horizon, T , and across all sensors. Stability is in the sense of the ℓ_1 -norm between the centralized and distributed versions of the posterior at each sensor, and at each time within T . Roughly speaking, our main result shows that uniform ε -stability of the distributed filtering process depends only loglinearly on T and (roughly) the size of the network, and only logarithmically on $1/\varepsilon$. If this total loglinear bound is fulfilled, any additional consensus iterations will incur a fully quantified further exponential decay in the consensus error. Our bounds are universal, in the sense that they are independent of the particular structure of the Gaussian Hidden Markov Model (HMM) under consideration.

Keywords. Nonlinear Filtering, Markov Chains, ADMM, Average Consensus, Distributed State Estimation, Hidden Markov Models.

1 Introduction

Distributed state estimation and tracking of partially observed processes constitute central problems in modern Distributed Networks of Agents (DNAs), where, in the absence of a powerful fusion center, a group of power constrained sensing devices with limited computation and/or communication capabilities receive noisy measurements of some common, possibly rapidly varying process of global network interest. As also surveyed in [2], important applications of DNAs, where potentially the aforementioned problems arise, include environmental and agricultural monitoring [3], health care monitoring [4], pollution source localization [5], surveillance [6], chemical plume tracking [7], target tracking [7] and habitat monitoring [8], to name a few.

In the context of linear state estimation, in recent years and in parallel with the rapid development of fast averaging consensus protocols and algorithms in networked measurement systems

The Authors are with the Department of Electrical & Computer Engineering, Rutgers, The State University of New Jersey, 94 Brett Rd, Piscataway, NJ 08854, USA. e-mail: {d.kalogerias, athinap}@rutgers.edu.

Part of this work [1] was presented at the *49th Asilomar Conference on Signals, Systems & Computers (Asilomar 2015)*. This work is supported by the National Science Foundation (NSF) under Grants CCF-1526908 & CNS-1239188.

[9–14], there has been extensive research on the important basic problem of distributed Kalman filtering over DNAs, under several different perspectives, such as utilizing control theoretic consensus algorithms [15–17], the sign-of-innovations approach [18], the Alternating Direction Method of Multipliers (ADMM) [19], which will also be considered in this work and others, based on more customized consensus strategies [20,21]. For a relatively complete list of references, see the extensive survey [22].

Although not as rich, the literature concerning the problem of distributed nonlinear filtering of general Markov processes is itself quite extensive. Since, in general, most nonlinear filters do not admit finite dimensional representations, the focus in this case is the derivation of distributed schemes for the implementation of well defined, finite dimensional nonlinear filtering approximations, which would allow for efficient, real time state estimation. In this direction, successful examples include distributed particle filters [2], distributed extended Kalman filters [16,18], as extensions to the linear case mentioned above and, more recently, the Bayesian consensus filtering approach proposed in [23].

In this paper, we focus on distributed state estimation of Markov chains with *finite state space*, partially observed in *conditionally Gaussian noise*. Hereafter, we will use the term Gaussian Hidden Markov Model (HMM). This special case is, however, of broad practical interest. Further, it is known that the posterior probability measure of the chain at a certain time, relative to the measurements obtained so far, admits a recursive representation [24,25]. This posterior can subsequently be used in order to produce any optimal estimate of interest, such as the MMSE estimator of the chain, the respective MAP estimator, etc. The noisy observations of the chain are obtained distributively over a DNA, whose connectivity pattern follows a connected Random Geometric Graph (RGG) [26,27].

Under this setting, we consider a distributed filtering scheme, which relies on the distributed evaluation of the likelihood part of the centralized posterior under consideration and is based on a particular specialization of the ADMM for fast average consensus [14]. Our idea is similar to the one conveyed by the respective formulation for particle filters [2]. In order to account for rapidly varying Markov chains with arbitrary statistical structure, in our formulation, nonlinear filtering and distributed average consensus are implemented in *different time scales*, with the message exchange rate between any sensor and its neighbors being much larger than the rate of measurement acquisition [16]. This is a realistic assumption in systems where measurements are acquired at relatively distant times; for example, every minute, hour or day, or even more dynamically, possibly at event triggered time instants or a sequence of stopping times. Additionally, in this paper we assume perfect communications among sensors. This is a reasonable assumption, provided, for instance, that the filtering process is implemented in a higher than the physical layer of the network under consideration. Our contributions are summarized as follows:

1) First, based on the matrix equivalent formulation of the ADMM for implementing average consensus and the respective convergence results presented in [14], we focus on optimizing the respective consensus error bound, with respect to two free parameters: a scalar $\epsilon > 0$ and the second largest eigenvalue of a symmetric (doubly) stochastic matrix \mathbf{S} ([14], also see Section 2). Under indeed mild conditions, we show analytically that the error bound is optimized (in a certain sense) at an explicitly defined ϵ and by choosing \mathbf{S} such that its second largest eigenvalue is minimized, the latter being a convex optimization problem, which can be efficiently solved [11]. Our results essentially complement earlier work on ADMM-based consensus, previously presented in [14].

2) Then, utilizing the aforementioned optimized bounds and assuming the same number of consensus iterations between any two consecutive measurements for each sensor in the network, we fully characterize a minimal number of such iterations required, such that the distributed filter

remains *uniformly* stable with a prescribed accuracy level, $\varepsilon \in (0, 1]$, within a finite operational horizon, T . Uniform stability is in the sense of the supremum of the ℓ_1 -norm between the centralized and distributed versions of the posterior in each sensor, over all sensors and over all times within the operational horizon of interest. Roughly speaking, our result shows that, under very reasonable assumptions, the stability of the distributed filtering process depends only *loglinearly* on T and the total number of measurements in the network, N , and only *logarithmically* on $1/\varepsilon$. Additionally, if this total loglinear bound is fulfilled, any additional consensus iterations will incur an *exponential decrease* in the aforementioned ℓ_1 -norm. The result is *fundamental* and *universal*, since, apart from the assumed conditional Gaussianity of the observations, it is virtually *independent of the internal structure* of the particular HMM under consideration. This fact makes it particularly attractive in highly heterogeneous DNAs.

The problem of distributed inference in HMMs has been considered earlier in [28]. However, the approach taken in [28] is distinctly different from our proposed approach. In particular, the formulation in [28] is based on control theoretic consensus [10] and stochastic approximation, while our distributed filtering formulation is based on optimized ADMM-based consensus. Also, the method of [28] requires certain assumptions on the statistical structure of the hidden Markov chain (e.g., primitiveness), while our work makes no such assumptions. Further, although the analysis presented in [28] does provide some limited convergence guarantees, it does not provide any results on the rate of convergence of the obtained distributed estimators. Here, we provide a fully tractable stability analysis of the distributed filtering scheme considered, with explicit, optimistic and universal bounds on the rate of convergence, as well as the degree of consensus achieved. Another important difference is that, in [28], filtering and consensus are implemented simultaneously. Such setting excludes cases where the underlying Markov chain is very rapidly varying and cannot fully exploit the benefits of heterogeneity in the information observed by the sensors of the DNA under consideration. This is due to the fact that, in [28], global consensus cannot, in general, be guaranteed within a reasonable and quantitatively predictable error margin. Contrary to [28], the distributed filtering scheme advocated herein efficiently exploits sensor heterogeneity, since uniform diffusion of local information is achieved across the network.

The paper is organized as follows. In Section II, we present the system model in detail, along with some very basic preliminaries on nonlinear filtering. Section III introduces our distributed filtering formulation and, subsequently, focuses on the optimization of the consensus error bounds produced by the ADMM iterations. In Section IV, we study stability of the distributed filtering scheme considered in the sense mentioned above, and we present the relevant results, along with complete proofs. In Section V, we present some numerical simulations, experimentally validating some of the properties of the proposed approach, as well as a relevant discussion. Finally, Section VI concludes the paper.

Notation: In the following, the state vector will be represented as X_t , and all other matrices and vectors will be denoted by boldface uppercase and boldface lowercase letters, respectively. Real valued random variables will be denoted by uppercase letters. Calligraphic letters and formal script letters will denote sets and σ -algebras, respectively. The operators $(\cdot)^T$, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ will denote transposition, minimum and maximum eigenvalue, respectively. The ℓ_p -norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is $\|\mathbf{x}\|_p \triangleq (\sum_{i=1}^n |x(i)|^p)^{1/p}$, for all naturals $p \geq 1$. For any Euclidean space, \mathbf{I} will denote the respective identity operator of appropriate dimension. For any set \mathcal{A} , its complement is denoted as \mathcal{A}^c . Additionally, we employ the identifications $\mathbb{N}^+ \equiv \{1, 2, \dots\}$, $\mathbb{N}_n^+ \equiv \{1, 2, \dots, n\}$, $\mathbb{N}_n \equiv \{0\} \cup \mathbb{N}_n^+$ and $\mathbb{N}_n^m \equiv \mathbb{N}_n^+ \setminus \mathbb{N}_{m-1}^+$, for any positive naturals $n > m$.

2 System Model & Preliminaries

In this section, we first present a generic model of the class of systems under consideration, where multiple, possibly indirectly connected sensors observe noisy versions of a common, hidden (i.e., unobserved), stochastically evolving underlying signal of interest. Second, we briefly discuss the centralized solution to problem of inference of the underlying signal on the basis of the observations at the sensors.

2.1 System Model: HMMs over DNAs

Without loss of generality, we consider a wireless network consisting of S sensors, located inside the fixed square geometric region $\mathcal{S} \triangleq [0, 1]^2$. The connectivity pattern of the DNA under consideration is assumed to obey an *undirected* RGG model [26, 27]. According to such a model, the positions of the sensors, $\mathbf{p}_i \in \mathcal{S}, i \in \mathbb{N}_S^+$, are chosen uniformly at random in \mathcal{S} , whereas two sensors i and j are considered connected if and only if $\|\mathbf{p}_i - \mathbf{p}_j\|_2 \leq r$, where $r \in (0, \sqrt{2}]$ denotes the connectivity threshold of the network. Throughout the paper, we assume that the underlying RGG of the network is *simply connected*, which constitutes an event happening with very high probability for sufficiently large number of sensors and/or sufficiently large connectivity threshold [27]. Also, the one hop neighborhood of each sensor i , including itself, is denoted as \mathbb{N}_i , for all $i \in \mathbb{N}_S^+$.

At each discrete time instant $t \in \mathbb{N}$, sensor i measures the vector process $\mathbf{y}_t^i \in \mathbb{R}^{N_i \times 1}$, for all $i \in \mathbb{N}_S^+$, where $\sum_{i \in \mathbb{N}_S^+} N_i \triangleq N$. It is assumed that the stacked process $\mathbf{y}_t \triangleq \left[\left(\mathbf{y}_t^1 \right)^T \dots \left(\mathbf{y}_t^S \right)^T \right]^T \in \mathbb{R}^{N \times 1}$ constitutes a noisy version (i.e., a functional) of a (for simplicity) time-homogeneous Markov chain X_t , called the *state*, with finite state space \mathcal{X} of cardinality $L \in \mathbb{N}^+$. Conditioned on X_t , the process \mathbf{y}_t is distributed according to a jointly Gaussian law as

$$\mathbf{y}_t | X_t \sim \mathcal{N}(\boldsymbol{\mu}_t(X_t), \boldsymbol{\Sigma}_t(X_t)), \quad (1)$$

where $\boldsymbol{\mu}_t : \mathcal{X} \mapsto \mathbb{R}^{N \times 1}$ and $\boldsymbol{\Sigma}_t : \mathcal{X} \mapsto \mathbb{R}^{N \times N}$ constitute known measurable functionals of X_t , for all $t \in \mathbb{N}$. In particular, $\boldsymbol{\Sigma}_t(\cdot)$ is assumed to be of block diagonal form, constituting of the submatrices $\mathbb{R}^{N_i \times N_i} \ni \boldsymbol{\Sigma}_t^i(\mathbf{x}) \succ \mathbf{0}$, for all $i \in \mathbb{N}_S^+$, also implying that $\boldsymbol{\Sigma}_t(\mathbf{x}) \succ \mathbf{0}$, for all $\mathbf{x} \in \mathcal{X}$. As a result, the measurements at each sensor are statistically independent, given X_t . Likewise, we define $\boldsymbol{\mu}_t(\mathbf{x}) \triangleq \left[\left(\boldsymbol{\mu}_t^1(\mathbf{x}) \right)^T \dots \left(\boldsymbol{\mu}_t^S(\mathbf{x}) \right)^T \right]^T$, where $\boldsymbol{\mu}_t^i(\mathbf{x}) \in \mathbb{R}^{N_i \times 1}$, for all $i \in \mathbb{N}_S^+$ and $\mathbf{x} \in \mathcal{X}$. Further, the following additional technical assumptions are made.

Assumption: (Boundedness) The quantities $\lambda_{\max}(\boldsymbol{\Sigma}_t(\mathbf{x}))$, $\|\boldsymbol{\mu}_t(\mathbf{x})\|_2$ are both uniformly upper bounded in $t \in \mathbb{N}$ and $\mathbf{x} \in \mathcal{X}$, with finite bounds λ_{\sup} and μ_{\sup} , respectively. For technical reasons, it is also true that $\lambda_{\inf} \triangleq \inf_{t \in \mathbb{N}} \inf_{\mathbf{x} \in \mathcal{X}} \lambda_{\min}(\boldsymbol{\Sigma}_t(\mathbf{x})) \geq e > 1$, a requirement which can always be satisfied by normalization of the observations.

Regarding the hidden Markov chain under consideration, it is true that $\mathcal{X} \triangleq \{\mathbf{x}_1, \dots, \mathbf{x}_L\}$, where \mathbf{x}_i denotes the i -th state of the chain, for $i \in \mathbb{N}_L^+$. Although it is actually irrelevant, for simplicity we will assume that $\mathbf{x}_i \in \mathbb{R}$, for all $i \in \mathbb{N}_L^+$. The temporal dynamics of X_t are expressed in terms of the initial distribution of the chain, encoded in the vector $\pi_{-1} \in \mathbb{R}^{L \times 1}$, as well as the *column stochastic* transition matrix $\mathbf{P} \in [0, 1]^{L \times L}$, defined, as usual, as $\mathbf{P}(i, j) \triangleq \mathcal{P}(X_t \equiv \mathbf{x}_i | X_{t-1} \equiv \mathbf{x}_j)$, for all $(i, j) \in \mathbb{N}_L^+ \times \mathbb{N}_L^+$.

2.2 Centralized Recursive Estimates

Let $\{\mathcal{Y}_t\}_{t \in \mathbb{N}}$ be the complete filtration generated by the observations \mathbf{y}_t . A quantity of central importance in nonlinear filtering is the posterior probability measure of the state X_t given \mathcal{Y}_t , which, since the state space of X_t is finite, can be represented by a random vector $\pi_{t|\mathcal{Y}_t} \in \mathbb{R}^{L \times 1}$. Under the system model defined above, this random vector can be exactly evaluated in real time as [24, 25]

$$\pi_{t|\mathcal{Y}_t} \equiv \frac{\mathbf{E}_t}{\|\mathbf{E}_t\|_1}, \quad \forall t \in \mathbb{N}, \quad (2)$$

where the process $\mathbf{E}_t \in \mathbb{R}^{L \times 1}$ satisfies the linear recursion $\mathbf{E}_t \equiv \mathbf{\Lambda}_t \mathbf{P} \mathbf{E}_{t-1}$, for all $t \in \mathbb{N}$, with $\mathbf{E}_{-1} \triangleq \pi_{-1}$,

$$\mathbf{\Lambda}_t \triangleq \text{diag}(\mathbf{L}_t(\mathbf{x}_1) \dots \mathbf{L}_t(\mathbf{x}_L)) \in \mathbb{R}^{L \times L}, \quad (3)$$

$$\mathbf{L}_t(\mathbf{x}) \triangleq \frac{\exp\left(-\frac{1}{2} \bar{\mathbf{y}}_t^T(\mathbf{x}) \mathbf{\Sigma}_t^{-1}(\mathbf{x}) \bar{\mathbf{y}}_t(\mathbf{x})\right)}{\sqrt{\det(\mathbf{\Sigma}_t(X_t))}} \in \mathbb{R}, \text{ and} \quad (4)$$

$$\bar{\mathbf{y}}_t(\mathbf{x}) \triangleq \mathbf{y}_t - \boldsymbol{\mu}_t(\mathbf{x}) \in \mathbb{R}^{N \times 1}, \quad \forall \mathbf{x} \in \mathcal{X} \text{ and } \forall t \in \mathbb{N}. \quad (5)$$

Of course, under these circumstances and defining $\mathbf{X} \triangleq [\mathbf{x}_1 \dots \mathbf{x}_L] \in \mathbb{R}^{1 \times L}$, the MMSE estimator of X_t given \mathcal{Y}_t can be easily obtained as $\hat{X}_t \triangleq \mathbb{E}\{X_t | \mathcal{Y}_t\} \equiv \mathbf{X} \pi_{t|\mathcal{Y}_t}$, for all $t \in \mathbb{N}$. Similar estimates may be obtained for other quantities of possible interest, such as the posterior error covariance matrix between X_t and \hat{X}_t , etc. In the following, we focus on the distributed estimation of the random probability measure encoded in $\pi_{t|\mathcal{Y}_t}$ (and not specific functionals of it).

Remark 1. The structure of the nonlinear filtering procedure outlined above may be interpreted as follows. Quantity (4) constitutes the *conditional likelihood* of the observations, *given* that the state (the hidden process) is fixed. As a result, the filtering recursion for \mathbf{E}_t may be decomposed into two intuitively informative parts, namely, a *one-step ahead prediction* step ($\mathbf{P} \mathbf{E}_{t-1}$), and a *correction/update* step $\mathbf{\Lambda}_t(\mathbf{P} \mathbf{E}_{t-1})$. Then, (re)normalization by $\|\mathbf{E}_t\|_1$ converts the produced estimate into a valid probability measure. ■

3 Distributed Estimation via the ADMM

Employing somewhat standard notation, used, for instance, in [14], exploiting the block-diagonal structure of $\mathbf{\Sigma}_t$ and, therefore, of $\mathbf{\Sigma}_t^{-1}$ as well, and with $\bar{\mathbf{y}}_t^i(\mathbf{x}) \triangleq \mathbf{y}_t^i - \boldsymbol{\mu}_t^i(\mathbf{x})$, we may reexpress (4) as

$$\begin{aligned} \mathbf{L}_t(\mathbf{x}) &\equiv \exp\left(-\frac{1}{2} \left(\bar{\mathbf{y}}_t^T(\mathbf{x}) \mathbf{\Sigma}_t^{-1}(\mathbf{x}) \bar{\mathbf{y}}_t(\mathbf{x}) + \log \det(\mathbf{\Sigma}_t(\mathbf{x})) \right)\right) \\ &\equiv \exp\left(-\frac{1}{2} \sum_{i \in \mathbb{N}_S^+} \left[\left(\bar{\mathbf{y}}_t^i(\mathbf{x}) \right)^T \left(\mathbf{\Sigma}_t^i(\mathbf{x}) \right)^{-1} \bar{\mathbf{y}}_t^i(\mathbf{x}) + \log \det(\mathbf{\Sigma}_t^i(\mathbf{x})) \right]\right), \end{aligned} \quad (6)$$

for all $\mathbf{x} \in \mathcal{X}$ and $t \in \mathbb{N}$. Let

$$\theta_t^i(\mathbf{x}) \triangleq S \left(\bar{\mathbf{y}}_t^i(\mathbf{x}) \right)^T \left(\mathbf{\Sigma}_t^i(\mathbf{x}) \right)^{-1} \bar{\mathbf{y}}_t^i(\mathbf{x}) + S \log \det \left(\mathbf{\Sigma}_t^i(\mathbf{x}) \right), \quad \forall i \in \mathbb{N}_S^+, \quad \forall \mathbf{x} \in \mathcal{X} \text{ and} \quad (7)$$

$\boldsymbol{\theta}_t(\mathbf{x}) \triangleq [\theta_t^1(\mathbf{x}) \dots \theta_t^S(\mathbf{x})]^T \in \mathbb{R}^{S \times 1}, \forall t \in \mathbb{N}$. Then, it is true that

$$\theta_t(\mathbf{x}) \triangleq \sum_{i \in \mathbb{N}_S^+} \frac{\theta_t^i(\mathbf{x})}{S} \equiv \bar{\mathbf{y}}_t^T(\mathbf{x}) \boldsymbol{\Sigma}_t^{-1}(\mathbf{x}) \bar{\mathbf{y}}_t(\mathbf{x}) + \log \det(\boldsymbol{\Sigma}_t(\mathbf{x})). \quad (8)$$

Of course, sensor i of the network observes θ_t^i , for all $t \in \mathbb{N}$. It is then clear that knowledge of θ_t at each sensor is equivalent to the ability of locally evaluating the posterior $\pi_{t|\mathcal{Y}_t}$, since $\exp(-\theta_t(\mathbf{x})/2) \equiv \mathbf{L}_t(\mathbf{x})$, which is the likelihood part of the filter at time t . There are exactly L possibilities for X_t and, as a result, in order to make the computation of the likelihood matrix $\boldsymbol{\Lambda}_t$ locally feasible, each sensor should be able to acquire $\mathbf{L}_t(\mathbf{x}_j)$, for all $j \in \mathbb{N}_L^+$. But since there are no functional dependencies among the latter quantities, each can be separately computed, in a completely parallel fashion. Consequently, it suffices to focus exclusively on $\mathbf{L}_t(\mathbf{x}_j)$, for some fixed $j \in \mathbb{N}_L^+$.

Freeze time t temporarily and let $\epsilon > 0$. Also, consider a *symmetric (doubly) stochastic* matrix $\mathbf{S} \in \mathbb{R}^{S \times S}$, with $\mathbf{S}(k, l)$ being non zero if and only if $l \in \mathbb{N}_k$, for all $k \in \mathbb{N}_S^+$; precise selection of \mathbf{S} will be considered later (see Theorem 3 and relevant discussion). Then, employing the ADMM based, distributed averaging *Method B* as discussed in [14], with $\epsilon \mathbf{S}$ as the matrix of augmentation constants [14] and exploiting its symmetricity, it can be easily shown that the ADMM iterations [14] for the distributed computation of the average (8), at sensor $k \in \mathbb{N}_S^+$, are reduced to

$$\boldsymbol{\vartheta}_t^j(n, k) \equiv \frac{1}{1 + \epsilon} \theta_t^k(\mathbf{x}_j) + \ell_t(n - 1, k) + \frac{\epsilon}{(1 + \epsilon)} \sum_{l \in \mathbb{N}_k} \mathbf{S}(k, l) \boldsymbol{\vartheta}_t^j(n - 1, l) \quad \text{and} \quad (9)$$

$$\ell_t(n, k) \equiv \ell_t(n - 1, k) + \frac{\epsilon}{2(1 + \epsilon)} \left(\sum_{l \in \mathbb{N}_k} \mathbf{S}(k, l) \boldsymbol{\vartheta}_t^j(n - 1, l) - \boldsymbol{\vartheta}_t^j(n - 1, k) \right), \quad (10)$$

for all $n \in \mathbb{N}^2$. The scheme is initialized as

$$\boldsymbol{\vartheta}_t^j(1, k) \triangleq \frac{1}{1 + \epsilon} \theta_t^k(\mathbf{x}_j) \quad \text{and} \quad (11)$$

$$\ell_t(1, k) \triangleq 0, \quad \forall k \in \mathbb{N}_S^+. \quad (12)$$

Now, letting t vary in \mathbb{N} , at each iteration n_t and at each sensor k , the true posterior measure $\pi_{t|\mathcal{Y}_t}$ is approximated as

$$\tilde{\pi}_{t|\mathcal{Y}_t}^k(n_t) \triangleq \frac{\tilde{\mathbf{E}}_t^k(n_t)}{\|\tilde{\mathbf{E}}_t^k(n_t)\|_1}, \quad \forall t \in \mathbb{N}, \quad (13)$$

where the process $\tilde{\mathbf{E}}_t^k(n_t) \in \mathbb{R}^{L \times 1}$ satisfies the linear recursion $\tilde{\mathbf{E}}_t^k(n_t) \equiv \tilde{\boldsymbol{\Lambda}}_t^k(n_t) \mathbf{P} \tilde{\mathbf{E}}_{t-1}^k(n_{t-1})$, for all $t \in \mathbb{N}$, with $\tilde{\mathbf{E}}_{-1}^k(n_t) \equiv \tilde{\mathbf{E}}_{-1}^k \triangleq \pi_{-1} \equiv \mathbf{E}_{-1}$ and where $\tilde{\boldsymbol{\Lambda}}_t^k(n_t)$ is defined as in (3), but with the likelihoods $\mathbf{L}_t(\mathbf{x}_j), j \in \mathbb{N}_L^+$ being replaced by

$$\tilde{\mathbf{L}}_t^{n_t, k}(\mathbf{x}_j) \triangleq \exp\left(-\frac{\boldsymbol{\vartheta}_t^j(n_t, k)}{2}\right), \quad \forall j \in \mathbb{N}_L^+, \quad (14)$$

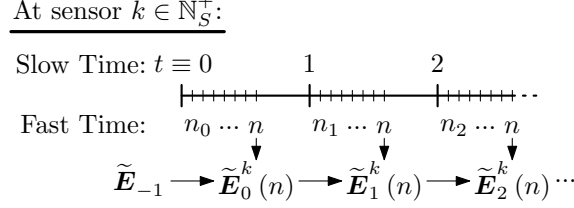


Figure 1: An intuitive schematic representation of the distributed filtering procedure at sensor $k \in \mathbb{N}_S^+$.

completing the algorithmic description of the distributed HMM under consideration. In the above, n_t denotes the *iteration index at time t* . In this paper, mainly for analytical and intuitional simplicity, we will assume that $n_t \equiv n$, for all $t \in \mathbb{N}$. This simply means that each estimate $\tilde{\mathbf{E}}_t^k(n)$ is recursively constructed using the “same-iteration” estimate $\tilde{\mathbf{E}}_{t-1}^k(n)$, at the previous time step. See Fig. 1 for a schematic representation of the procedure explained above, for each sensor in the DNA under consideration.

Remark 2. For practical considerations, at this point, it would be worth specifying both the computational complexity and the information exchange complexity of the distributed filtering considered, per sensor $k \in \mathbb{N}_S^+$, per slow time $t \in \mathbb{N}$. With regards to computational complexity, it is easy to see that, at sensor k , each consensus iteration incurs an order of $\mathcal{O}(L|\mathbf{N}_k|)$ operations of combined multiply-adds. Therefore, multiplying by the total number of consensus iterations per slow time, n (the fast time), and adding the computational complexity incurred by filtering, results in total computational complexity at sensor k , for each slow time, of the order of $\mathcal{O}(nL|\mathbf{N}_k| + L^2)$. One can see that, for a hidden chain with state space of large cardinality, L , if $n|\mathbf{N}_k| \approx L$, then the computational complexity of the distributed filtering scheme under consideration is of the same order as that of filtering alone. As far as information exchange complexity is concerned, following a similar procedure, it can be readily shown that an order of $\mathcal{O}(n|\mathbf{N}_k|)$ bidirectional (at worst) message exchanges have to take place at sensor k , per slow time t . ■

Arguing as in [14], it is then straightforward to show that of central importance concerning the convergence of the ADMM scheme is the eigenstructure and, in particular, the Second Largest Eigenvalue Modulus (SLEM) [29] of the matrix

$$\mathbf{M} \triangleq \begin{bmatrix} \frac{\epsilon}{1+\epsilon} \mathbf{S} + \mathbf{I} & -\frac{\epsilon}{2(1+\epsilon)} (\mathbf{S} + \mathbf{I}) \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{2S \times 2S}. \quad (15)$$

Hereafter, let the sets $\{\lambda_{\mathbf{S}}^i\}_{i \in \mathbb{N}_S^+}$ and $\{\lambda_{\mathbf{M}}^i\}_{i \in \mathbb{N}_{2S}^+}$ contain the eigenvalues of \mathbf{S} and \mathbf{M} , respectively. Then, recalling that, due to symmetricity, the spectrum of \mathbf{S} is real, set $\lambda_{\mathbf{S}}^1 \triangleq 1$ and $\lambda_{\mathbf{S}}^2 \triangleq \max_{i \in \{2, \dots, S\}} \lambda_{\mathbf{S}}^i$. Because the zero pattern of \mathbf{S} is the same as that of the adjacency matrix of the connected RGG modeling the connectivity of the DNA under consideration (with ones in its diagonal entries), it can be easily shown that \mathbf{S} corresponds to the transition matrix of a Markov chain, which is both irreducible and aperiodic. Therefore, the SLEM of \mathbf{S} is strictly smaller than one, and, consequently, $\lambda_{\mathbf{S}}^2 < 1$ as well [29]. Likewise, since \mathbf{M} has a unique unit eigenvalue [14], let $\lambda_{\mathbf{M}}^1 \equiv |\lambda_{\mathbf{M}}^1| \triangleq 1$, and also let $0 < \rho < 1$ (also see [14]) denote the SLEM of \mathbf{M} , that is, $\rho \triangleq \max_{i \in \{2, \dots, 2S\}} |\lambda_{\mathbf{M}}^i|$. For later reference, define the auxiliary vector

$\boldsymbol{\vartheta}_t^j(n) \triangleq [\boldsymbol{\vartheta}_t^j(n, 1) \dots \boldsymbol{\vartheta}_t^j(n, S)]^T \in \mathbb{R}^{S \times 1}$. When applied to the average consensus problem, and specifically to the distributed filtering problem under consideration, the rate of convergence of the ADMM is characterized as follows.

Theorem 1. (ADMM Consensus Error Bound [14]) *Fix a natural $T > 0$, denoting a finite operational horizon. For each $t \in \mathbb{N}_T^+$ and an arbitrary $j \in \mathbb{N}_L^+$, it is true that, for all $n \in \mathbb{N}^2 \cap \mathbb{N}^{[\epsilon+1]}$,*

$$\left\| \boldsymbol{\vartheta}_t^j(n) - \boldsymbol{\vartheta}_t^j(\infty) \right\|_2 \leq (1 + \epsilon)^{-1} \left\| \boldsymbol{\theta}_t(\mathbf{x}_j) \right\|_2 n \rho^{n-1}, \quad (16)$$

where $\boldsymbol{\vartheta}_t^j(\infty) \triangleq \boldsymbol{\theta}_t(\mathbf{x}_j) \mathbf{1}_S$.

An apparent conclusion of Theorem 1 is that the rate of convergence of the distributed averaging procedure depends to a large extent on ϵ and of course the SLEM of \mathbf{M} , ρ , which, by construction, is also a function of ϵ . On the other hand, ρ depends on the eigenstructure of \mathbf{S} , also by construction. As it turns out, the full spectrum of \mathbf{M} is explicitly related to the spectrum of \mathbf{S} . The relevant result follows, complementing the analysis previously presented in [14].

Theorem 2. (Characterization of the Spectrum of \mathbf{M}) *Given $\lambda_{\mathbf{S}}^i, i \in \mathbb{N}_S^+$ and for any $\epsilon > 0$, the eigenvalues of \mathbf{M} may be expressed in pairs as*

$$\lambda_{\mathbf{M}}^{2i-1, 2i}(\lambda_{\mathbf{S}}^i, \epsilon) \equiv \frac{1 + \epsilon + \epsilon \lambda_{\mathbf{S}}^i \pm \sqrt{1 + \epsilon^2 \left((\lambda_{\mathbf{S}}^i)^2 - 1 \right)}}{2(1 + \epsilon)}, \quad (17)$$

for all $i \in \mathbb{N}_S^+$. Then, the SLEM of \mathbf{M} can be expressed as

$$\rho(\epsilon, \lambda_{\mathbf{S}}^2) \equiv \begin{cases} \left| \lambda_{\mathbf{M}}^3(\lambda_{\mathbf{S}}^2, \epsilon) \right|, & \text{if } \epsilon \leq h(\lambda_{\mathbf{S}}^2) \\ \frac{\epsilon}{1 + \epsilon}, & \text{otherwise} \end{cases}, \quad \forall \epsilon > 0, \quad (18)$$

where, for all $\lambda_{\mathbf{S}}^2 \in [-1, 1)$,

$$h(\lambda_{\mathbf{S}}^2) \triangleq \frac{1 + \lambda_{\mathbf{S}}^2}{1 - \lambda_{\mathbf{S}}^2} \mathbf{1}_{[0, 1)}(\lambda_{\mathbf{S}}^2) + \mathbf{1}_{[-1, 0)}(\lambda_{\mathbf{S}}^2). \quad (19)$$

Further, for fixed $\lambda_{\mathbf{S}}^2 \in [-1, 1)$, ρ is globally minimized at

$$\epsilon^*(\lambda_{\mathbf{S}}^2) \triangleq \begin{cases} \frac{1}{\sqrt{1 - (\lambda_{\mathbf{S}}^2)^2}}, & \text{if } \lambda_{\mathbf{S}}^2 \in [0, 1) \\ 1, & \text{if } \lambda_{\mathbf{S}}^2 \in [-1, 0) \end{cases}, \quad (20)$$

with optimal value

$$\rho^*(\lambda_{\mathbf{S}}^2) \equiv \begin{cases} \frac{\lambda_{\mathbf{S}}^2 + 1 + \sqrt{1 - (\lambda_{\mathbf{S}}^2)^2}}{2 \left(1 + \sqrt{1 - (\lambda_{\mathbf{S}}^2)^2} \right)}, & \text{if } \lambda_{\mathbf{S}}^2 \in [0, 1) \\ 1/2, & \text{if } \lambda_{\mathbf{S}}^2 \in [-1, 0) \end{cases}, \quad (21)$$

constituting a continuous and increasing function of the second largest eigenvalue of \mathbf{S} , $\lambda_{\mathbf{S}}^2$.

Proof of Theorem 2. In principle, certain parts of Theorem 2 (in particular (17)) may be derived based on the analysis presented in [14]. However, the derivation is not straightforward. For the reader's convenience, a self contained proof of Theorem 2 is presented in Appendix A. ■

It will be important to note that all results presented so far regarding convergence of the ADMM based consensus schema, as well as the eigenstructure of \mathbf{M} , hold equally well even if we relax our obvious demand that the elements of the symmetric (doubly) stochastic matrix \mathbf{S} , corresponding to the non zero entries of the adjacency matrix of the RGG under consideration, are strictly positive. Such an assertion can be validated by inspecting the respective proof of Theorem 1 in [14] and observing that the aforementioned assumption on the entries \mathbf{S} is actually not required; it is the eigenstructure of \mathbf{S} that really matters. On the other hand, the property of \mathbf{S} having its second largest eigenvalue strictly smaller than unity is not destroyed by relaxing strict positivity of its elements; the relaxation corresponds to an enlargement of the space of possible choices for \mathbf{S} [29]. In addition to the above, it can be readily verified by Theorem 2 that, if $\lambda_{\mathbf{S}}^2 < 1$, \mathbf{M} will have exactly one unit eigenvalue and that, under the same condition, $\rho(\epsilon, \lambda_{\mathbf{S}}^2) < 1$, satisfying the requirement for convergence implied by Theorem 1.

All the above will be critical for us, since if we assume that the diagonal entries of \mathbf{S} may be nonnegative, then there are increased degrees of freedom for choosing the best \mathbf{S} , such that the consensus error bound of Theorem 1 is optimized, and there are well established tools for achieving this goal [11], as we will discuss below. However, it should be mentioned that mere nonnegativity of the entries of \mathbf{S} implies that the consensus scheme employed for distributed filtering *does not constitute an ordinary ADMM variant anymore*, since in the usual ADMM based consensus formulation, the constants multiplying the quadratic part of the augmented Lagrangian are positive by default. This fact though does not compromise performance; in fact, it results in a faster distributed averaging scheme. Therefore, hereafter, we will assume that the zero pattern of \mathbf{S} “includes” that of the adjacency matrix of the underlying RGG, *in the sense that all elements of \mathbf{S} are allowed to be nonnegative, except for those corresponding to the zero pattern of the aforementioned adjacency matrix.*

Fig. 2 shows the graph of the optimal SLEM $\rho^*(\lambda_{\mathbf{S}}^2)$ as a function of $\lambda_{\mathbf{S}}^2$, as described in Theorem 2, as well as the nonoptimized SLEM $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$, for a set of fixed suboptimal choices of ϵ 's. The obvious fact that $\rho^*(\lambda_{\mathbf{S}}^2)$ is a convex, non affine and increasing function of $\lambda_{\mathbf{S}}^2$ is very important, because it implies that $\rho^*(\lambda_{\mathbf{S}}^2)$ can be further optimized with respect to the matrix \mathbf{S} , resulting in substantial performance gain as $\lambda_{\mathbf{S}}^2$ moves towards zero. In particular, it is true that

$$\min_{\mathbf{S} \in \mathfrak{S}} \min_{\epsilon \in \mathbb{R}_{++}} \rho(\epsilon, \lambda_{\mathbf{S}}^2) \equiv \min_{\mathbf{S} \in \mathfrak{S}} \rho^*(\lambda_{\mathbf{S}}^2) \equiv \rho^*\left(\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2\right), \quad (22)$$

possibly for different decisions on \mathbf{S} on the center and right of (22), where \mathfrak{S} denotes the feasible set of symmetric (doubly) stochastic matrices, with $\mathbf{S}(k, l)$ being zero if $l \notin \mathbf{N}_k$, for all $k \in \mathbf{N}_S^+$. It is well known that the optimization problem on the right of (22) is convex and that it can be solved efficiently either via semidefinite programming, or subgradient methods [11]. Therefore, the objective on the left of (22) can be globally optimized as well.

Quite surprisingly, under some additional, albeit very mild, conditions on the ADMM iteration index, n , optimization of the SLEM of $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$, either with respect to ϵ or both ϵ and \mathbf{S} indeed

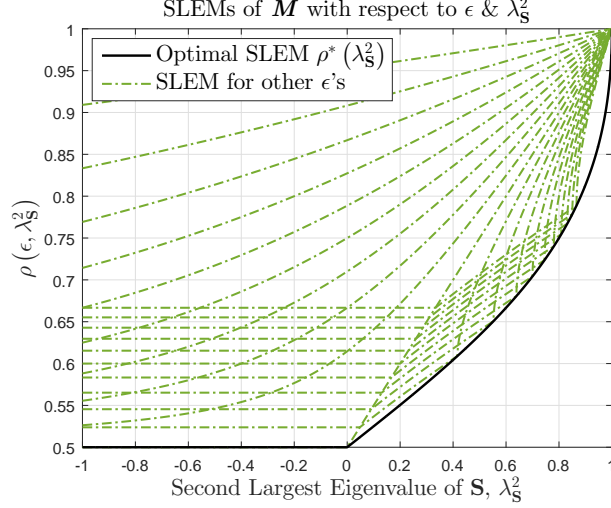


Figure 2: SLEMs of \mathbf{M} as a function of the largest eigenvalue of \mathbf{S} , $\lambda_{\mathbf{S}}^2$, for various fixed values of ϵ (green dash-dot line), as well as the optimal choice $\epsilon^*(\lambda_{\mathbf{S}}^2)$ (black solid line).

implies optimization of the consensus error bound of the ADMM, as stated in Theorem 1, under a certain sense. In this respect, we present the next result.

Theorem 3. (ADMM Optimal Consensus Error Bound I) *Fix a natural $T > 0$ and choose $\epsilon_{max} \geq \epsilon^*(\lambda_{\mathbf{S}}^2)$, $\lambda_{\mathbf{S}}^2 \in [-1, 1)$, for some $\mathbf{S} \in \mathfrak{S}$. For each $t \in \mathbb{N}_T^+$ and any $j \in \mathbb{N}_L^+$, as long as $n > 2\epsilon_{max} + 1$, the choice $\epsilon \equiv \epsilon^*(\lambda_{\mathbf{S}}^2)$ is Pareto optimal for the scalar, multiobjective program*

$$\underset{\epsilon > 0}{\text{minimize}} \left[\left\{ \left(\rho(\epsilon, \lambda_{\mathbf{S}}^2) \right)^{n-1} / (1 + \epsilon) \text{ s.t. } \epsilon < n \right\}_n \right], \quad (23)$$

optimizing the RHS of (16), resulting in the optimal consensus error bound

$$\left\| \boldsymbol{\vartheta}_t^j(n) - \boldsymbol{\vartheta}_t^j(\infty) \right\|_2 \leq \gamma(\lambda_{\mathbf{S}}^2) \left\| \boldsymbol{\theta}_t(\mathbf{x}_j) \right\|_2 n \left(\rho^*(\lambda_{\mathbf{S}}^2) \right)^n, \quad (24)$$

where

$$\gamma(\lambda_{\mathbf{S}}^2) \triangleq \begin{cases} \frac{2\sqrt{1 - (\lambda_{\mathbf{S}}^2)^2}}{1 + \lambda_{\mathbf{S}}^2 + \sqrt{1 - (\lambda_{\mathbf{S}}^2)^2}}, & \text{if } \lambda_{\mathbf{S}}^2 \in [0, 1) \\ 1, & \text{if } \lambda_{\mathbf{S}}^2 \in [-1, 0) \end{cases}, \quad (25)$$

being a continuous and decreasing function of $\lambda_{\mathbf{S}}^2$ in $[-1, 1)$. Additionally, choose any $\tilde{\mathbf{S}} \in \mathfrak{S}$. If, under the same setting as above, $\lambda_{\mathbf{S}}^2$ is replaced by the optimal value $\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2 \leq \lambda_{\tilde{\mathbf{S}}}^2$ in (24), then the resulting bound is optimal, in the sense that

$$\min_{\substack{\mathbf{S} \in \mathfrak{S} \\ \lambda_{\mathbf{S}}^2 \in [-1, \lambda_{\tilde{\mathbf{S}}}^2]}} \gamma(\lambda_{\mathbf{S}}^2) \left(\rho^*(\lambda_{\mathbf{S}}^2) \right)^n \equiv \gamma\left(\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2\right) \left(\rho^*\left(\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2\right) \right)^n. \quad (26)$$

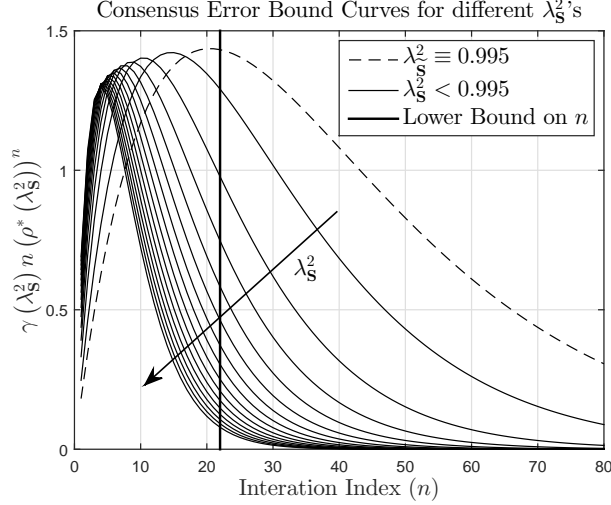


Figure 3: Demonstration of the second part of Theorem 3: Consensus error bounds for various values of $\lambda_{\tilde{\mathbf{S}}}^2$, when $\tilde{\mathbf{S}}$ is chosen such that $\lambda_{\tilde{\mathbf{S}}}^2 \equiv 0.995$.

Proof of Theorem 3. See Appendix B. ■

A graphical demonstration of the second, somewhat technical part of Theorem 3 is shown in Fig. 3. As it can be readily observed, given an arbitrary, possibly “bad” $\tilde{\mathbf{S}} \in \mathfrak{S}$, corresponding to a specific, possibly large, second largest eigenvalue $\lambda_{\tilde{\mathbf{S}}}^2$ (in the figure, this value equals 0.995), the best convergence rate is attained at $\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2$, when the feasible set is constrained such that $\lambda_{\mathbf{S}}^2 \in [-1, \lambda_{\tilde{\mathbf{S}}}^2]$. Also note that, because $\epsilon^*(\lambda_{\tilde{\mathbf{S}}}^2)$ is increasing in $\lambda_{\tilde{\mathbf{S}}}^2$, Pareto efficiency of (24) is preserved when $\lambda_{\tilde{\mathbf{S}}}^2$ is replaced by $\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2$.

Some of the technicalities of Theorem 3 may be avoided if one focuses on optimizing the looser consensus error bound

$$\left\| \boldsymbol{\vartheta}_t^j(n) - \boldsymbol{\vartheta}_t^j(\infty) \right\|_2 \leq \left\| \boldsymbol{\theta}_t(\mathbf{x}_j) \right\|_2 n \rho^{n-1}, \quad (27)$$

since $(1 + \epsilon)^{-1} < 1$. Indeed, the following result holds; proof is omitted.

Theorem 4. (ADMM Optimal Consensus Error Bound II) Fix a natural $T > 0$ and choose $\epsilon_{max} \geq \epsilon^*(\lambda_{\tilde{\mathbf{S}}}^2)$, $\lambda_{\tilde{\mathbf{S}}}^2 \in [-1, 1)$, for some $\tilde{\mathbf{S}} \in \mathfrak{S}$. For each $t \in \mathbb{N}_T^+$ and any $j \in \mathbb{N}_L^+$, the globally optimal consensus error bound corresponding to the RHS of (27), with respect to

$$(\epsilon, \mathbf{S}) \in (0, \epsilon_{max}] \times \left\{ \mathbf{C} \in \mathfrak{S} \mid \lambda_{\mathbf{C}}^2 \in [-1, \lambda_{\tilde{\mathbf{S}}}^2] \right\}, \quad (28)$$

is given by

$$\left\| \boldsymbol{\vartheta}_t^j(n) - \boldsymbol{\vartheta}_t^j(\infty) \right\|_2 \leq \left\| \boldsymbol{\theta}_t(\mathbf{x}_j) \right\|_2 n \left(\rho^* \left(\min_{\mathbf{S} \in \mathfrak{S}} \lambda_{\mathbf{S}}^2 \right) \right)^{n-1}, \quad (29)$$

for all $n > \max \{2, \epsilon_{max}\}$.

Finally, in our main results, presented in Section 4, the quantity $\tau \triangleq 1/\log(1/\rho)$ will be of great importance in establishing a minimal number of ADMM iterations required between each pair of subsequent observation times, ensuring stability of the distributed nonlinear filter under consideration, over a finite horizon of filtering operation, T . Interestingly, τ does not appear coincidentally

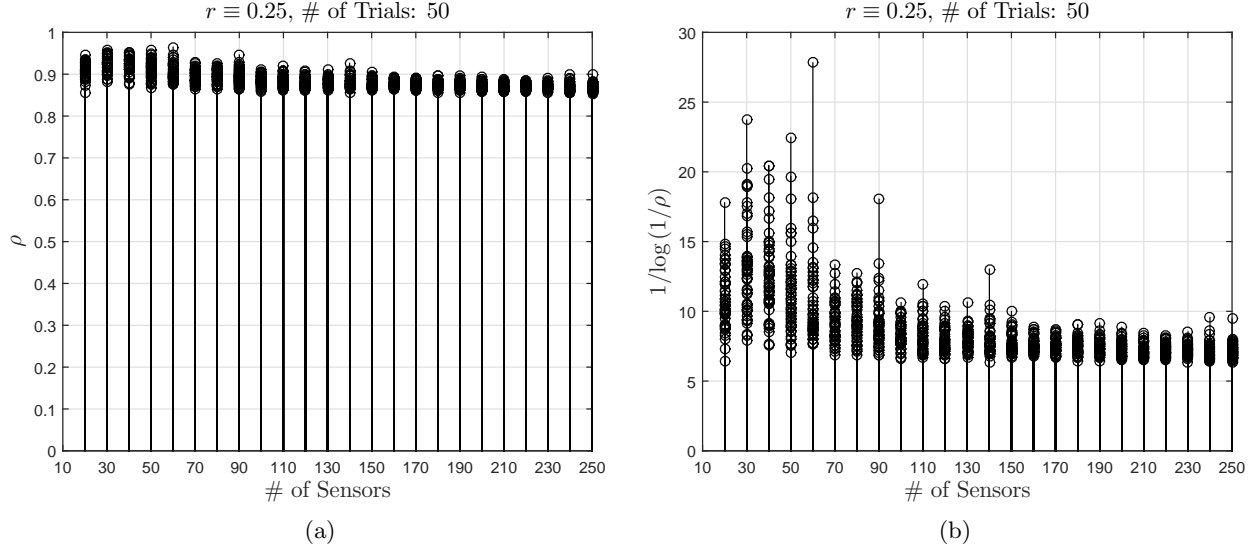


Figure 4: The quantities (a) ρ and (b) $\tau \equiv 1/\log(1/\rho)$, as functions of the number of sensors S . For each value of S , 50 independent realizations of the respective quantities are shown.

in our analysis; when analyzing the mixing properties of Markov chains over graphs, τ constitutes a known quantity, called the *mixing time* of the chain [29]. In this fashion, we call τ the mixing time of \mathbf{M} , characterizing the convergence rate of the ADMM.

Of course, it is both expected and intuitively correct that ρ and, therefore, τ , will be both dependent on the connectivity density of the underlying RGG (measured, for instance, through the sparsity of the adjacency matrix of the graph), modeling the connectivity of the DNA under study. More specifically, we should at least expect that ρ (which constitutes a random variable) will be in general smaller in “more strongly” connected RGGs, and the same for τ . Also, since more sensors uniformly scattered in the same area mean essentially higher connectivity density, the same behavior is expected as the number of sensors increase. For example, consider the case where we are given a DNA corresponding to a *strongly* connected RGG. Unfortunately though, an analytic characterization of the aforementioned properties boils down to analyzing the relevant properties of $\lambda_{\mathbf{S}}^2$, which constitutes a notoriously difficult (and still open) mathematical problem. Despite of this difficulty, we can verify the aforementioned behavior experimentally. Indeed, Fig. 4 shows τ (a) and ρ (b) as functions of the number of sensors, S . For each value of S , 50 trials are presented. In this example, \mathbf{S} is suboptimally chosen according to the *maximum degree chain* construction [29], whereas ϵ is optimally chosen. We readily observe that our intuitive assertions stated above are successfully verified. Based on these results, in all the subsequent analysis, we will consider τ and ρ to be *bounded and in general decreasing* functions of the number of sensors in the network, S , and/or the connectivity threshold, r . This fact constitutes the direct conclusion of Fig. 4. It is not straightforward to determine how the actual performance of the whole distributed filtering scheme, is affected by S , the threshold r and the corresponding values of ρ and τ . This will be precisely the topic of interest in the next section, where we show that, in order to guarantee ϵ -global agreement on filtering estimates, it is sufficient that the number of consensus iterations per sensor, per slow time, n , scales linearly with τ and, at worst, loglinearly with S (note that, on average, S is inversely proportional to τ , as discussed above).

4 Uniform Stability of Distributed Filtering

This section is devoted to the presentation and proof of the main results of the paper. Hereafter, for notational brevity, we will set $\gamma(\lambda_S^2) \equiv \gamma$ and $\rho^*(\lambda_S^2) \equiv \rho$, in accordance to Theorem 3 presented in Section 3. Therefore, it will also be assumed that everything holds as long as $n > 2\epsilon_{max} + 1$, on top of any other condition on the iteration index, n . If, for any reason, one wishes to consider Theorems 1 or 4, the aforementioned statements have to be modified accordingly.

The following lemmata, borrowed from [30], will be helpful in the subsequent analysis.

Lemma 1. (Telescoping Bound for Matrix Products [30]) *Consider the collections of arbitrary, square matrices*

$$\left\{ \mathbf{A}_i \in \mathbb{C}^{N \times N} \right\}_{i \in \mathbb{N}_n} \quad \text{and} \quad \left\{ \mathbf{B}_i \in \mathbb{C}^{N \times N} \right\}_{i \in \mathbb{N}_n}.$$

For any submultiplicative matrix norm $\|\cdot\|_{\mathfrak{M}}$, it is true that¹

$$\left\| \prod_{i=0}^n \mathbf{A}_i - \prod_{i=0}^n \mathbf{B}_i \right\|_{\mathfrak{M}} \leq \sum_{i=0}^n \left(\prod_{j=0}^{i-1} \|\mathbf{A}_j\|_{\mathfrak{M}} \right) \left(\prod_{j=i+1}^n \|\mathbf{B}_j\|_{\mathfrak{M}} \right) \|\mathbf{A}_i - \mathbf{B}_i\|_{\mathfrak{M}}. \quad (30)$$

Lemma 2. (Stability of Observations [30]) *Consider the random quadratic form*

$$Q_t(\omega) \triangleq \|\mathbf{y}_t(\omega)\|_2^2 \equiv \|\bar{\mathbf{y}}_t(X_t(\omega)) + \boldsymbol{\mu}_t(X_t(\omega))\|_2^2, \quad t \in \mathbb{N}. \quad (31)$$

Then, for any fixed $t \in \mathbb{N}$ and any freely chosen $C \geq 1$, there exists a bounded constant $\beta > 1$, such that the measurable set

$$\mathcal{T}_t \triangleq \left\{ \omega \in \Omega \left| \sup_{i \in \mathbb{N}_t} Q_i(\omega) < \beta C N (1 + \log(t+1)) \right. \right\} \quad (32)$$

satisfies

$$\mathcal{P}(\mathcal{T}_t) \geq 1 - (t+1)^{1-CN} \exp(-CN), \quad (33)$$

that is, the sequence of quadratic forms $\{Q_i(\omega)\}_{i \in \mathbb{N}_t}$ is uniformly bounded with overwhelmingly high probability, under the probability measure \mathcal{P} .

The next result constitutes a corollary of Theorem 1. Here, for our purposes, we state it as a preliminary lemma.

Lemma 3. (Rate of each Iterate) *For each $t \in \mathbb{N}_T^+$ and an arbitrary $j \in \mathbb{N}_L^+$, it is true that, for all $n \in \mathbb{N}^{[2\epsilon_{max}+2]}$,*

$$\left| \boldsymbol{\vartheta}_t^j(n, k) - \boldsymbol{\theta}_t(\mathbf{x}_j) \right| \leq \gamma \|\boldsymbol{\theta}_t(\mathbf{x}_j)\|_2 n \rho^n, \quad (34)$$

for all $k \in \mathbb{N}_S^+$.

Proof of Lemma 3. See Appendix C. ■

¹Hereafter, we adopt the conventions $\prod_{j=0}^{-1}(\cdot) \equiv \prod_{j=n+1}^n(\cdot) \triangleq 1$.

Employing Lemma 2, we can characterize the growth of the the initial vector $\boldsymbol{\theta}_t(\mathbf{x}_j)$, for $t \in \mathbb{N}_T$, as follows.

Lemma 4. (Growth of Initial Values) *There exists a $\delta > 1$, such that*

$$\sup_{t \in \mathbb{N}_T^+} \sup_{j \in \mathbb{N}_L^+} \|\boldsymbol{\theta}_t(\mathbf{x}_j)\|_2 \leq \delta S^{1.5} C N (1 + \log(T+1)), \quad (35)$$

with probability at least $1 - (T+1)^{1-CN} \exp(-CN)$, for any free $C \geq 1$.

Proof of Lemma 4. See Appendix D. ■

Let us now present another lemma, providing a probabilistic uniform lower bound concerning the iterates $\boldsymbol{\vartheta}_t^j(n, k)$, under appropriate conditions on the number of iterations, n .

Lemma 5. (A lower bound on each Iterate) *It is true that, as long as $n \in \mathbb{N}^{\lfloor 2\epsilon_{max}+2 \rfloor}$ and*

$$n - \tau \log(\gamma n) \geq \tau \log \left(\frac{2\delta S^{1.5} C (1 + \log(T+1))}{\log(\lambda_{inf})} \right), \quad (36)$$

the iterates generated by the ADMM are lower bounded as

$$\inf_{t \in \mathbb{N}_T^+} \inf_{j \in \mathbb{N}_L^+} \inf_{k \in \mathbb{N}_S^+} \boldsymbol{\vartheta}_t^j(n, k) \geq \frac{N}{2} \log(\lambda_{inf}), \quad (37)$$

with probability at least $1 - (T+1)^{1-CN} \exp(-CN)$.

Proof of Lemma 5. See Appendix E. ■

Leveraging the results presented above, we may now provide a probabilistic bound on the required number of ADMM iterations, such that the ℓ_1 norm between the **unnormalized** filtering estimates \mathbf{E}_t and $\tilde{\mathbf{E}}_t^k(n)$ is small.

Theorem 5. (Stability of Unnormalized Distributed Estimates) *Fix $T < \infty$ and choose any global accuracy level $0 < \varepsilon < 1$. Then, there exists $\eta > 1$ such that, as long as $n \in \mathbb{N}^{\lfloor 2\epsilon_{max}+2 \rfloor}$ and*

$$n - \tau \log(\gamma n) \geq \tau \log \left(\frac{\eta S^{1.5} C (1 + \log(T+1))}{\varepsilon} \right), \quad (38)$$

the absolute error between \mathbf{E}_t and $\tilde{\mathbf{E}}_t^k(n)$ satisfies

$$\sup_{t \in \mathbb{N}_T} \sup_{k \in \mathbb{N}_S^+} \left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 \leq \varepsilon, \quad (39)$$

with probability at least $1 - (T+1)^{1-CN} \exp(-CN)$. That is, provided (38) holds locally at each sensor $k \in \mathbb{N}_S^+$, at each $t \in \mathbb{N}_T$, \mathbf{E}_t equals $\tilde{\mathbf{E}}_t^k(n)$ **within** ε , with overwhelmingly high probability.

Proof of Theorem 5. First, we know that $\mathbf{E}_t \equiv \mathbf{\Lambda}_t \mathbf{P} \mathbf{E}_{t-1}$, for all $t \in \mathbb{N}$, where $\mathbf{E}_{-1} \equiv \pi_{-1}$. Therefore, simple induction shows that

$$\mathbf{E}_t = \left(\prod_{i \in \mathbb{N}_t} (\mathbf{\Lambda}_{t-i} \mathbf{P}) \right) \mathbf{E}_{-1}, \quad \forall t \in \mathbb{N}. \quad (40)$$

Likewise, by construction of our distributed state estimator,

$$\tilde{\mathbf{E}}_t^k(n) = \left(\prod_{i \in \mathbb{N}_t} (\tilde{\mathbf{\Lambda}}_{t-i}^k(n) \mathbf{P}) \right) \tilde{\mathbf{E}}_{-1}^k, \quad \forall t \in \mathbb{N} \quad (41)$$

and for all $k \in \mathbb{N}_S^+$, where $\tilde{\mathbf{E}}_{-1}^k \equiv \mathbf{E}_{-1}$, identically. Taking the difference between \mathbf{E}_t and $\tilde{\mathbf{E}}_t^k(n)$, we can write

$$\mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) = \left(\prod_{i \in \mathbb{N}_t} (\mathbf{\Lambda}_{t-i} \mathbf{P}) - \prod_{i \in \mathbb{N}_t} (\tilde{\mathbf{\Lambda}}_{t-i}^k(n) \mathbf{P}) \right) \mathbf{E}_{-1}. \quad (42)$$

Then, taking the ℓ_1 -norm on both sides of (42), using the fact that $\|\mathbf{E}_{-1}\|_1 \equiv 1$ and invoking Lemma 1, we get

$$\begin{aligned} & \left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 \\ & \leq \sum_{i \in \mathbb{N}_t} \left(\prod_{j=0}^{i-1} \|\mathbf{\Lambda}_{t-j}\|_1 \|\mathbf{P}\|_1 \right) \left(\prod_{j=i+1}^t \|\tilde{\mathbf{\Lambda}}_{t-j}^k(n)\|_1 \|\mathbf{P}\|_1 \right) \left\| (\mathbf{\Lambda}_{t-i} - \tilde{\mathbf{\Lambda}}_{t-i}^k(n)) \right\|_1 \|\mathbf{P}\|_1. \end{aligned} \quad (43)$$

Of course, the ℓ_1 -norm for a matrix (induced by the usual ℓ_1 -norm for vectors) coincides with its maximum column sum. As a result, it will be true that $\|\mathbf{P}\|_1 \equiv 1$, since \mathbf{P} constitutes a column stochastic matrix. Now, recalling the internal structure of the diagonal matrices $\mathbf{\Lambda}_t$ and $\tilde{\mathbf{\Lambda}}_t^k$, $t \in \mathbb{N}$, the fact that λ_{inf} is strictly greater than unity and known apriori and Lemma 5 stated and proved above, it will be true that

$$\|\mathbf{\Lambda}_{t-i}\|_1 \leq \frac{1}{\lambda_{inf}^{N/2}} \quad \text{and} \quad \|\tilde{\mathbf{\Lambda}}_{t-i}^k(n)\|_1 \leq \frac{1}{\lambda_{inf}^{N/4}}, \quad (44)$$

for all $i \in \mathbb{N}_t$, under the conditions the aforementioned lemma suggests. Then, the RHS of (43) can be further bounded as

$$\begin{aligned} \left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 & \leq \sum_{i \in \mathbb{N}_t} \frac{\left\| (\mathbf{\Lambda}_{t-i} - \tilde{\mathbf{\Lambda}}_{t-i}^k(n)) \right\|_1}{\lambda_{inf}^{Ni/2} \lambda_{inf}^{N(t-i)/4}} \\ & \equiv \frac{1}{\lambda_{inf}^{Nt/4}} \sum_{i \in \mathbb{N}_t} \frac{\left\| (\mathbf{\Lambda}_{t-i} - \tilde{\mathbf{\Lambda}}_{t-i}^k(n)) \right\|_1}{\lambda_{inf}^{Ni/4}}. \end{aligned} \quad (45)$$

Next, focusing on each term inside the summation above, we have (see Lemma 4)

$$\begin{aligned}
\left\| \left(\mathbf{\Lambda}_{t-i} - \tilde{\mathbf{\Lambda}}_{t-i}^k(n) \right) \right\|_1 &\equiv \max_{l \in \mathbb{N}_L^+} \left| \exp \left(-\frac{1}{2} \theta_{t-i}(\mathbf{x}_j) \right) - \exp \left(-\frac{1}{2} \boldsymbol{\vartheta}_{t-i}^l(n, k) \right) \right| \\
&\leq \max_{l \in \mathbb{N}_L^+} \left| \theta_{t-i}(\mathbf{x}_j) - \boldsymbol{\vartheta}_{t-i}^l(n, k) \right| \leq \max_{l \in \mathbb{N}_L^+} \gamma \left\| \boldsymbol{\theta}_{t-i}(\mathbf{x}_j) \right\|_2 n \rho^n \\
&\leq \gamma \delta S^{1.5} C N (1 + \log(T+1)) n \rho^n.
\end{aligned} \tag{46}$$

Using the above into (45), we arrive at the inequality

$$\begin{aligned}
\left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 &\leq \frac{\gamma \delta S^{1.5} C N (1 + \log(T+1)) n \rho^n}{\lambda_{inf}^{Nt/4}} \sum_{i \in \mathbb{N}_t} \frac{1}{\lambda_{inf}^{Ni/4}} \\
&\leq \frac{\gamma \delta S^{1.5} C N (1 + \log(T+1)) n \rho^n}{\lambda_{inf}^{Nt/4}} \left(t + \frac{1}{\lambda_{inf}^{N/4}} \right) \\
&\leq \gamma \delta S^{1.5} C (1 + \log(T+1)) n \rho^n \left(\frac{Nt}{\lambda_{inf}^{Nt/4}} + \frac{N(t+1)}{\lambda_{inf}^{N(t+1)/4}} \right) \\
&\leq \frac{8\gamma \delta S^{1.5} C (1 + \log(T+1)) n \rho^n}{\log(\lambda_{inf}) \lambda_{inf}^{1/\log(\lambda_{inf})}}.
\end{aligned} \tag{47}$$

But because $\lambda_{inf}^{1/\log(\lambda_{inf})} \equiv e$ and $(\log(\lambda_{inf}))^{-1} \leq 1$, we get

$$\left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 \leq 3\gamma \delta S^{1.5} C (1 + \log(T+1)) n \rho^n, \tag{48}$$

holding true for all $t \in \mathbb{N}_T$, with probability at least $1 - (T+1)^{1-CN} \exp(-CN)$, provided that

$$n - \tau \log(\gamma n) \geq \tau \log \left(\frac{2\delta S^{1.5} C (1 + \log(T+1))}{\log(\lambda_{inf})} \right). \tag{49}$$

Finally, choose an $\varepsilon \in (0, 1)$. Then, a sufficient condition such that $\left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 \leq \varepsilon$, for all $t \in \mathbb{N}_T$ and $k \in \mathbb{N}_S^+$, is

$$3\gamma \delta S^{1.5} C (1 + \log(T+1)) n \rho^n \leq \varepsilon, \tag{50}$$

or, after some algebra,

$$n - \tau \log(\gamma n) \geq \tau \log \left(\frac{3\delta S^{1.5} C (1 + \log(T+1))}{\varepsilon} \right). \tag{51}$$

Of course, the bounds (49) and (51) must hold at the same time. Thus, for each $t \in \mathbb{N}_T$, $\varepsilon \in (0, 1)$ and choosing

$$\eta \triangleq \delta \max \{ 2/\log(\lambda_{inf}), 3 \} > 1, \tag{52}$$

it will be true that $\sup_{t \in \mathbb{N}_T} \sup_{k \in \mathbb{N}_S^+} \left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1 \leq \varepsilon$, with probability at least $1 - (T+1)^{1-CN} \exp(-CN)$, provided that (38) holds, completing the proof. \blacksquare

Our main result follows, establishing a fundamental lower bound on the minimal number of consensus steps, such that the distributed filter remains *uniformly* stable with a prescribed accuracy level, within finite operational horizon and across all sensors in the network.

Theorem 6. (Stability of the Distributed HMM Estimator) *Fix a natural $1 \leq T < \infty$ and choose $\varepsilon \in (0, 1]$, $m \geq 0$, $C \geq 1$, and a number of sensors $S \geq 2$. Then, there exists a constant $c > 0$ such that, as long as $n \in \mathbb{N}^{\lfloor 2\varepsilon_{max} + 2 \rfloor}$ and*

$$\begin{aligned} n - \tau \log(\gamma n) &\geq c\tau CNT \log\left(\frac{CST}{\varepsilon}\right) + \tau m \\ &\equiv \mathcal{O}\left(\tau CNT \log\left(\frac{CST}{\varepsilon}\right)\right) + \tau m, \end{aligned} \quad (53)$$

the absolute error between $\pi_{t|\mathcal{Y}_t}$ and $\tilde{\pi}_{t|\mathcal{Y}_t}^k(n)$ satisfies

$$\sup_{k \in \mathbb{N}_S^+} \sup_{t \in \mathbb{N}_T} \left\| \pi_{t|\mathcal{Y}_t} - \tilde{\pi}_{t|\mathcal{Y}_t}^k(n) \right\|_1 \leq \varepsilon \exp(-m), \quad (54)$$

with probability at least $1 - (T + 1)^{1-CN} \exp(-CN)$. That is, provided (53) holds locally at each sensor $k \in \mathbb{N}_S^+$, $\pi_{t|\mathcal{Y}_t}$ equals $\tilde{\pi}_{t|\mathcal{Y}_t}^k(n)$ **within a global, exponentially decreasing factor** $\varepsilon \exp(-m)$, with overwhelmingly high probability.

Proof of Theorem 6. First, by construction and invoking the reverse triangle inequality, it is easy to show that the error in the ℓ_1 -norm between $\pi_{t|\mathcal{Y}_t}$ and $\tilde{\pi}_{t|\mathcal{Y}_t}^k(n)$ may be upper bounded as

$$\left\| \pi_{t|\mathcal{Y}_t} - \tilde{\pi}_{t|\mathcal{Y}_t}^k(n) \right\|_1 \leq 2 \frac{\left\| \mathbf{E}_t - \tilde{\mathbf{E}}_t^k(n) \right\|_1}{\left\| \mathbf{E}_t \right\|_1}, \quad (55)$$

for all $k \in \mathbb{N}_S^+$, $t \in \mathbb{N}_T$ and any qualifying $n \in \mathbb{N}^2$. In light of Theorem 5, as long as the respective conditions on n are fulfilled, it will be true that, for any $\varepsilon \in (0, 1)$,

$$\left\| \pi_{t|\mathcal{Y}_t} - \tilde{\pi}_{t|\mathcal{Y}_t}^k(n) \right\|_1(\omega) \leq \frac{2\varepsilon}{\left\| \mathbf{E}_t \right\|_1(\omega)}, \quad (56)$$

for all $\omega \in \mathcal{T}_T$, with $\mathcal{P}(\mathcal{T}_T) \geq 1 - (T + 1)^{1-CN} \exp(-CN)$. Now, as far as $\left\| \mathbf{E}_t \right\|_1$ is concerned, it has been shown by the authors that [30, 31]

$$\left\| \mathbf{E}_t \right\|_1(\omega) \geq \lambda_{sup}^{-N(T+1)/2} \exp\left(-\frac{\left(\sqrt{\beta CN(1+\log(T+1))} + \mu_{sup}\right)^2 (T+1)}{2\lambda_{inf}}\right) \triangleq B, \quad (57)$$

for all ω in the same measurable set \mathcal{T}_T , where all our probabilistic statements made so far have taken place in. As a result, and given that $\max\{\lambda_{sup}, e\} \equiv \lambda_{sup}$ by assumption, there exists a constant $c_0 > 0$, such that

$$\frac{1}{B} \lambda_{sup}^{-c_0 CNT \log(T)} \leq 1. \quad (58)$$

In order to show this, note that

$$\begin{aligned}
\frac{(T+1)}{2} \left(N + \frac{(\sqrt{\beta CN(1+\log(T+1))} + \mu_{sup})^2}{\lambda_{inf}} \right) &\leq \frac{(T+1)}{2} \left(N + \frac{\beta CN(1+\log(T+1)) (\mu_{sup}+1)^2}{\lambda_{inf}} \right) \\
&\leq (T+1) CN(1+\log(T+1)) \left(\frac{1}{2} + \frac{\beta (\mu_{sup}+1)^2}{2\lambda_{inf}} \right) \\
&\triangleq c_1 (T+1) CN(1+\log(T+1)) \\
&\leq c_1 c_2 TCN \log(T) \triangleq c_0 TCN \log(T), \tag{59}
\end{aligned}$$

whenever $T \geq 1$, for some positive constants c_1 and c_2 . Consequently, if, for each $t \in \mathbb{N}_T$, one chooses

$$\varepsilon \equiv \frac{1}{2} \lambda_{sup}^{-c_0 TCN \log(T)} \tilde{\varepsilon} \exp(-m) \leq 1, \tag{60}$$

for some $\tilde{\varepsilon} \in (0, 1]$ and $m \geq 0$, it will be true that

$$\sup_{k \in \mathbb{N}_S^+} \sup_{t \in \mathbb{N}_T} \left\| \pi_{t|\mathcal{Y}_t} - \tilde{\pi}_{t|\mathcal{Y}_t}^k(n) \right\|_1 \leq \tilde{\varepsilon} \exp(-m), \tag{61}$$

provided that

$$n - \tau \log(\gamma n) \geq \tau \log \left(\frac{\eta S^{1.5} C (1 + \log(T+1))}{\varepsilon} \right). \tag{62}$$

Replacing ε with the expression in (60) and for $S \geq 2$, the logarithmic term above may be upper bounded as

$$\begin{aligned}
\log \left(\frac{\eta S^{1.5} C (1 + \log(T+1))}{\varepsilon} \right) &\equiv \log \left(\frac{2\eta S^{1.5} C (1 + \log(T+1))}{\tilde{\varepsilon}} \lambda_{sup}^{c_0 TCN \log(T)} \exp(m) \right) \\
&\leq m + c_0 \log(\lambda_{sup}) CN T \log(T) + \log \left(\frac{2c_2 \eta S^{1.5} C \log(T)}{\tilde{\varepsilon}} \right) \\
&\triangleq m + \tilde{c}_0 CN T \log(T) + \log \left(\frac{2c_2 \eta S^{1.5} C \log(T)}{\tilde{\varepsilon}} \right) \\
&\leq m + \tilde{c}_0 CN T \log(T) + \tilde{c}_2 \log \left(\frac{CST}{\tilde{\varepsilon}} \right) \\
&\leq m + (\tilde{c}_0 + \tilde{c}_2) CN T \log \left(\frac{CST}{\tilde{\varepsilon}} \right) \\
&\triangleq m + c CN T \log \left(\frac{CST}{\tilde{\varepsilon}} \right), \tag{63}
\end{aligned}$$

for some other positive constant \tilde{c}_2 . Putting it altogether and renaming $\tilde{\varepsilon}$ to ε , we have finally shown that, for any $\varepsilon \in (0, 1]$, $m \geq 0$, $S \geq 2$ and any natural $1 \leq T < \infty$, there exists a constant $c > 0$, such that whenever

$$n - \tau \log(\gamma n) \geq c \tau CN T \log \left(\frac{CST}{\varepsilon} \right) + \tau m, \tag{64}$$

it is true that

$$\sup_{k \in \mathbb{N}_S^+} \sup_{t \in \mathbb{N}_T} \left\| \pi_{t|\mathcal{Y}_t} - \tilde{\pi}_{t|\mathcal{Y}_t}^k(n) \right\|_1 \leq \varepsilon \exp(-m), \quad (65)$$

with probability at least $1 - (T + 1)^{1-CN} \exp(-CN)$, and the proof is complete. \blacksquare

Remark 3. The linear-plus-logarithmic form of the LHS of (53) does not seem to have a specific physical interpretation, related to the conclusions of Theorem 6. In particular, the logarithmic term appearing on the LHS of (53) results as an artifact of our analytical development. Nevertheless, it is true that, for relatively small values of τ and γ and for sufficiently large n , the logarithmic term $\tau \log(\gamma n)$ is insignificant, compared to n , and may be considered relatively constant, that is, $n - \tau \log(\gamma n) \approx n - \tau \kappa(\gamma)$, for some κ , possibly dependent on γ , but independent of n . Therefore, the logarithmic artifact on the LHS of (53) may be heuristically considered as a positive, constant bias on the RHS as n increases, which, however, does not affect the asymptotic behavior of the result. \blacksquare

5 Discussion & Some Numerical Simulations

In this Section, we present some numerical simulations, experimentally validating some of the properties of the distributed filtering scheme considered in this paper, as well as a relevant discussion.

For our numerical simulations, we assume that we are given a DNA with $S \equiv 60$ sensors. The connectivity threshold of the underlying RGG r is set to 0.2, representing a relatively sparsely connected network. Each sensor in the network observes noisy versions of a Markov chain with $L \equiv 4$ states, state space $\mathcal{X} \equiv \{0.7, 1/2, 1, 1/3\}$, a stochastic transition matrix

$$\mathbf{P} \equiv \begin{bmatrix} 0.4 & 0.25 & 0.2 & 0.3 \\ 0.3 & 0.25 & 0.3 & 0.3 \\ 0.2 & 0.25 & 0.3 & 0.2 \\ 0.1 & 0.25 & 0.2 & 0.2 \end{bmatrix} \in \mathbb{R}^{4 \times 4}, \quad (66)$$

and initial distribution $\pi_{-1} \equiv [1\ 0\ 0\ 0]^T$. Regarding the observation model of the HMM under consideration, we assume that $N_k \equiv 2$, $\boldsymbol{\mu}_t^k(\mathbf{x}) \equiv \boldsymbol{\mu}(\mathbf{x}) \equiv -[\sin(\mathbf{x}) \sin(\mathbf{x})]^T$ and $\boldsymbol{\Sigma}_t^k(\mathbf{x}) \equiv \boldsymbol{\Sigma}^k$, with $\boldsymbol{\Sigma}^k(i, j) \equiv 2 \exp(-2|i - j|)$, for all $(i, j) \in \mathbb{N}_2^+ \times \mathbb{N}_2^+$ and for all $k \in \mathbb{N}_S^+$. As far as the distributed averaging scheme is concerned, \mathbf{S} is suboptimally chosen according to the maximum degree chain construction and ϵ is optimally chosen, according to what was previously stated in Section 3.

As a potential practical example for the model setting considered above, sensors could be deployed in an industrial facility, monitoring the hidden state (a mode) of a composite chemical process, at various locations inside the facility, at certain time intervals. In order to increase the information diversity at each location, multiple simultaneous measurements may be recorded, resulting in a correlated observation model, such as the one considered in our simulations. The distributed filtering scheme considered in this paper could then be employed, in order to perform global state estimation without the need of a dedicated fusion center.

As we discussed in Section 4, Theorem 6 provides a *sufficient* condition on the number of consensus iterations, such that the ℓ_1 consensus error between the centralized and distributed versions of the posterior measure $\pi_{t|\mathcal{Y}_t}$ is *exponentially small, uniformly* through the whole operational horizon of interest and across sensors, at the same time.

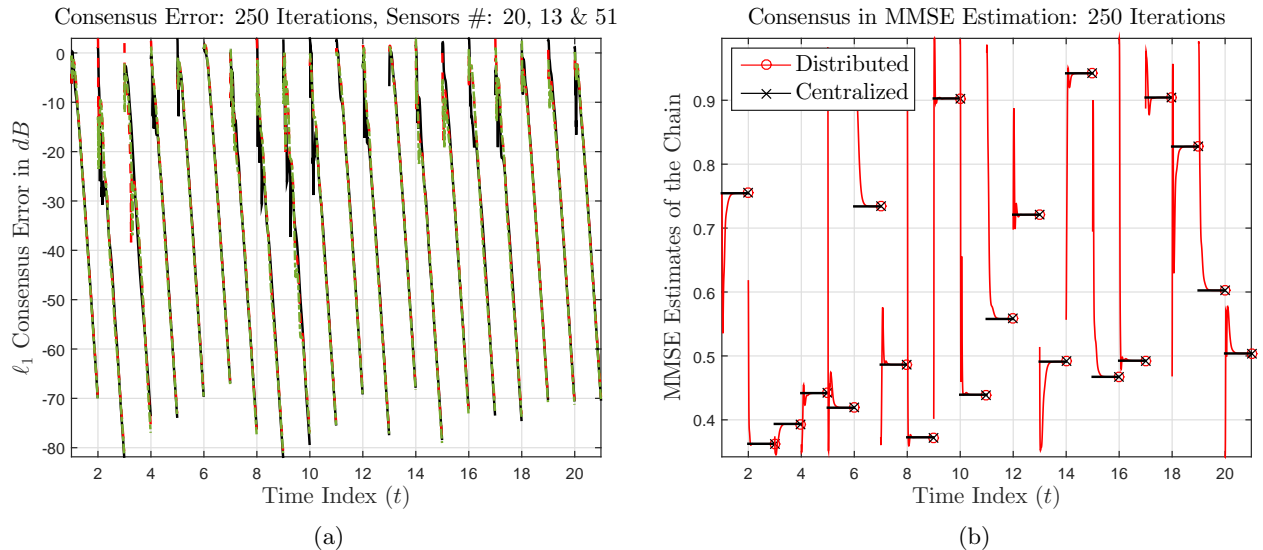


Figure 5: Case 2: 250 consensus iterations for $T \equiv 20$. (a) ℓ_1 consensus error curves. (b) Consensus in MMSE estimation. Between each pair of times, we plot the evolution of the respective distributed estimate.

Fig. 5a shows the ℓ_1 consensus error for $T \equiv 20$, where the number of iterations is set to 250. Results are shown for three randomly chosen sensors in the network. It is important to note that the number of iterations has been chosen such that, between each pair of consecutive times, the consensus error *does not reach machine precision*. This is critical in order to assess the uniform properties of the consensus error, through time and across sensors, and in order to verify the claims of our main result, Theorem 6. Indeed, as depicted in Fig. 5a, one can readily identify a perfectly acceptable uniform upper bound for the consensus error, for all 20 times of interest, say -60 dB. Additionally and in perfect agreement with Theorem 6, for the chosen number of consensus iterations, convergence of the averaging scheme is indeed *eventually* exponential, meaning that after some specific number of iterations, such type of convergence is indeed achievable. This is due to the fact that the consensus error curves appear roughly as straight lines in the logarithmic domain, a fact which translates to an exponential decay in the linear domain, respectively. The effect of this behavior on estimation quality can also be observed in Fig. 5b, which shows the MMSE estimates of the chain obtained by exploiting the respective distributed posterior estimates.

However, it is important to note that Theorem 6 does not provide a necessary condition on the behavior of the consensus error. This means that, within the framework of this paper, stability of distributed filtering is not theoretically predictable, in the case where condition (53) is not satisfied. In fact, for almost all numerical simulations we have conducted, the proposed scheme performed exceptionally well in terms of stability, even for relatively smaller number of iterations, at least as far as the given experimental setting is concerned. In particular, the error due to the imperfect agreement on the version of the posterior at each sensor does not significantly accumulate, as time progresses.

It is worth mentioning that, although loglinear dependence on $N \geq S$ and T of the required number of consensus iterations is not really evident from our numerical experiments, this is apparently due to the relatively small number of sensors and the relatively “easy” nonlinear filtering

problem considered. The aforementioned *asymptotic* loglinear behavior would certainly be revealed when the number of sensors and/or complexity of the HMM under consideration increases.

Finally, we would like to note that the method advocated in this paper is expected to perform better than the one in [28], since it utilizes more consensus iterations (fast time) per filtering time instant (slow time); [28] employs a single information exchange among the sensors, per filtering time instant. In [28], the filtering estimates at each sensor do not, in general, coincide within reasonable bounds, that is, consensus is not always achieved. In fact, consensus will not be achieved unless the network is *strongly connected*. On the other hand, in the distributed filtering scheme analyzed in this work, the information collected at each sensor is diffused uniformly through the whole network. As a result, consensus is achieved and any potential heterogeneity among the observations at the sensors is efficiently exploited.

6 Conclusion

We have studied stability of distributed inference in Gaussian HMMs with finite state space, exploiting ADMM-like based distributed average consensus. We have considered a distributed filtering scheme, which relies on the distributed evaluation of the likelihood part of the centralized nonlinear estimator under consideration. Basically assuming the same number of consensus iterations between any two consecutive sensor observations, we have provided a complete characterization of a minimal number of consensus steps, guaranteeing uniform stability of the resulting distributed nonlinear filter, within a finite operational horizon and across all sensors in the DNA under consideration. We have presented a fundamental result, showing that ε -stability of the distributed filtering process depends only loglinearly on the horizon of interest, T , and the dimensionality of the observation process, N , and logarithmically on $1/\varepsilon$. Moreover, strictly fulfilling this total loglinear bound incurs a fully quantified exponential decay in the consensus error. Our bounds are universal, in the sense that they are independent of the structure of the Gaussian HMM under consideration.

Appendix A: Proof of Theorem 2

As implied by the statement of Theorem 2, we start by studying the dependence of the spectrum of \mathbf{M} on that of \mathbf{S} and on $\epsilon > 0$. Of course, in order to determine the spectrum of \mathbf{M} , we look for the $2S$ solutions of the equation $\det(\mathbf{M} - \lambda\mathbf{I}) \equiv 0$, where

$$\mathbf{M} - \lambda\mathbf{I} \equiv \begin{bmatrix} \frac{\epsilon}{1+\epsilon}\mathbf{S} + (1-\lambda)\mathbf{I} & -\frac{\epsilon}{2(1+\epsilon)}(\mathbf{S} + \mathbf{I}) \\ \mathbf{I} & -\lambda\mathbf{I} \end{bmatrix}. \quad (67)$$

Since the submatrices \mathbf{I} and $-\lambda\mathbf{I}$ obviously commute in (67), we may invoke ([32], Theorem 3) and write

$$\begin{aligned} \det(\mathbf{M} - \lambda\mathbf{I}) &\equiv \det\left(-\frac{\epsilon\lambda}{1+\epsilon}\mathbf{S} - \lambda(1-\lambda)\mathbf{I} + \frac{\epsilon}{2(1+\epsilon)}(\mathbf{S} + \mathbf{I})\right) \\ &\equiv \frac{(-1)^S}{(2(1+\epsilon))^S} \det(\epsilon(2\lambda-1)\mathbf{S} + (2\lambda(1+\epsilon)(1-\lambda) - \epsilon)\mathbf{I}). \end{aligned} \quad (68)$$

Using (68), after some algebra and via a diagonalization of the symmetric \mathbf{S} , the equation $\det(\mathbf{M} - \lambda \mathbf{I}) \equiv 0$ can be easily shown to be equivalent to S quadratic equations of the form

$$-2(1 + \epsilon)\lambda^2 + 2(1 + \epsilon + \epsilon\lambda_{\mathbf{S}})\lambda - \epsilon(1 + \lambda_{\mathbf{S}}) \equiv 0, \quad (69)$$

where, for notational brevity, $\lambda_{\mathbf{S}}$ hereafter denotes any of the S eigenvalues comprising the spectrum of \mathbf{S} . Solving (69) results in the functional relation

$$\lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) \equiv \frac{1 + \epsilon + \epsilon\lambda_{\mathbf{S}} \pm \sqrt{1 + \epsilon^2((\lambda_{\mathbf{S}})^2 - 1)}}{2(1 + \epsilon)}, \quad (70)$$

for all $\lambda_{\mathbf{S}} \in (-1, 1]$ and $\epsilon > 0$, corresponding to the respective expression shown in the statement of Theorem 2.

The rest of the proof deals with characterizing (70) as a function of $\lambda_{\mathbf{S}}$ and ϵ , in order to determine the SLEM of \mathbf{M} , $\rho \equiv \rho(\epsilon, \lambda_{\mathbf{S}}^1, \dots, \lambda_{\mathbf{S}}^S)$ and, further, studying the problem of minimizing that SLEM with respect to ϵ . Fix $\epsilon > 0$. When $\lambda_{\mathbf{S}} \equiv 1$, then $\lambda_{\mathbf{M}}^+(1, \epsilon) \equiv 1 \equiv \lambda_{\mathbf{M}}^1$ and $\lambda_{\mathbf{M}}^-(1, \epsilon) \equiv \epsilon/(1 + \epsilon)$. This implies that the SLEM we are looking for is either $\epsilon/(1 + \epsilon)$, or is given by some $|\lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon)|$, for some $\lambda_{\mathbf{S}} \neq 1$. Thus, the SLEM of \mathbf{M} may be expressed as

$$\begin{aligned} \rho(\epsilon, \lambda_{\mathbf{S}}^2, \dots, \lambda_{\mathbf{S}}^S) &\equiv \max \left\{ \frac{\epsilon}{1 + \epsilon}, \left\{ \left| \lambda_{\mathbf{M}}^j(\lambda_{\mathbf{S}}^i, \epsilon) \right| \right\}_{i \in \mathbb{N}_S^2}^{j \in \{+, -\}} \right\} \\ &\equiv \max \left\{ \frac{\epsilon}{1 + \epsilon}, \max_{\lambda_{\mathbf{S}} \in \{\lambda_{\mathbf{S}}^i\}_{i \in \mathbb{N}_S^2}} \max_{j \in \{+, -\}} \left| \lambda_{\mathbf{M}}^j(\lambda_{\mathbf{S}}, \epsilon) \right| \right\}. \end{aligned} \quad (71)$$

Let us focus on the inner maximization on the RHS of (71). In the following, the cases where $\lambda_{\mathbf{S}}^2 \geq 0$ and $\lambda_{\mathbf{S}}^2 < 0$ will be treated separately. As we will see, the SLEM of \mathbf{M} behaves quite differently under each of the two aforementioned cases.

Suppose that $\lambda_{\mathbf{S}}^2 \geq 0$. Due to the existence of ϵ in the square root on the RHS of (70), it is reasonable to consider the following further two subcases:

- $1 + \epsilon^2((\lambda_{\mathbf{S}})^2 - 1) \geq 0$: If the condition on the left is true, then the corresponding eigenvalues of \mathbf{M} , compactly expressed via (70), are real. What is more, via an equivalence test, it can be easily shown that $\lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) > 0$ (for both nonnegative and negative $\lambda_{\mathbf{S}}$, actually). Consequently, in this case, $|\lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon)| \equiv \lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon)$, for each fixed ϵ . Under these considerations, because $\lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) > \lambda_{\mathbf{M}}^-(\lambda_{\mathbf{S}}, \epsilon)$ for any feasible $\lambda_{\mathbf{S}}$, as well as due to the easily provable fact that, whenever $\lambda_{\mathbf{S}} \geq 0$ and $\tilde{\lambda}_{\mathbf{S}} < 0$, $\lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) > \lambda_{\mathbf{M}}^+(\tilde{\lambda}_{\mathbf{S}}, \epsilon)$ (again via an equivalence test), it suffices to consider *maximizing* $\lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon)$ *only for the case of a nonnegative $\lambda_{\mathbf{S}}$* . Further, assuming that $\lambda_{\mathbf{S}} \geq 0$ (always in addition to the condition $1 + \epsilon^2((\lambda_{\mathbf{S}})^2 - 1) \geq 0$), it is easy to show that $\partial \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) / \partial \lambda_{\mathbf{S}} > 0$, implying that $\lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon)$ is *strictly increasing in $\lambda_{\mathbf{S}}$* , for any $\epsilon > 0$. Lastly, it is easy to show that $\lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) > \epsilon/(1 + \epsilon)$, regardless of $\lambda_{\mathbf{S}} \geq 0$ (a fact which will be useful later).

- $1 + \epsilon^2((\lambda_{\mathbf{S}})^2 - 1) < 0$: In this case, the respective pairs of eigenvalues of \mathbf{M} are complex.

Hence, it is true that

$$\left| \lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) \right| \equiv \frac{\sqrt{(1 + \epsilon + \epsilon \lambda_{\mathbf{S}})^2 - 1 - \epsilon^2 \left((\lambda_{\mathbf{S}})^2 - 1 \right)}}{2(1 + \epsilon)}. \quad (72)$$

By inspection, we have $\partial \left| \lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) \right| / \partial \lambda_{\mathbf{S}} > 0$. This again implies that $\left| \lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) \right|$ is *strictly increasing in $\lambda_{\mathbf{S}}$* , regardless of $\lambda_{\mathbf{S}}$, for any feasible choice of $\epsilon > 0$. Concerning comparison with $\epsilon / (1 + \epsilon)$, looking at $\left| \lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) \right|$ as a function of ϵ , it is true that

$$\left| \lambda_{\mathbf{M}}^{\pm}(\lambda_{\mathbf{S}}, \epsilon) \right| \geq \frac{\epsilon}{1 + \epsilon} \Leftrightarrow \epsilon \leq \frac{1 + \lambda_{\mathbf{S}}}{1 - \lambda_{\mathbf{S}}}, \quad (73)$$

which however holds true only for $\lambda_{\mathbf{S}} \geq 0$, since then and only then the inequality $(1 + \lambda_{\mathbf{S}}) / (1 - \lambda_{\mathbf{S}}) \geq 1 / \sqrt{1 - (\lambda_{\mathbf{S}})^2}$ is satisfied, where the RHS constitutes the lower bound for ϵ , due to the condition $1 + \epsilon^2 \left((\lambda_{\mathbf{S}})^2 - 1 \right) < 0$.

Now, for brevity, define the finite sets $\mathcal{C} \triangleq \mathcal{R}^c$, with

$$\mathcal{R} \triangleq \left\{ \lambda_{\mathbf{S}} \in \left\{ \lambda_{\mathbf{S}}^i \right\}_{i \in \mathbb{N}_{\mathbf{S}}^2} \mid 1 + \epsilon^2 \left((\lambda_{\mathbf{S}})^2 - 1 \right) \geq 0 \right\}, \quad (74)$$

both depending on the particular choice of $\epsilon > 0$. Then, from (71) and combining the arguments made previously, the SLEM under study may be expressed as

$$\rho \left(\epsilon, \lambda_{\mathbf{S}}^2, \dots, \lambda_{\mathbf{S}}^S \right) \equiv \max \left\{ \frac{\epsilon}{1 + \epsilon}, \max \left\{ \max_{0 \leq \lambda_{\mathbf{S}} \in \mathcal{R}} \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon), \max_{\lambda_{\mathbf{S}} \in \mathcal{C}} \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) \right| \right\} \right\}. \quad (75)$$

Apparently, there are two possibilities for $\lambda_{\mathbf{S}}^2$; either $\lambda_{\mathbf{S}}^2 \in \mathcal{C}$, or $\lambda_{\mathbf{S}}^2 \in \mathcal{R}$. If $\lambda_{\mathbf{S}}^2 \in \mathcal{C}$ and due to the monotonicity of the involved eigenvalues with respect to $\lambda_{\mathbf{S}}$, it is true that $\max_{\lambda_{\mathbf{S}} \in \mathcal{C}} \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) \right| \equiv \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^2, \epsilon) \right|$. Also, if there is any chance for the problem $\max_{0 \leq \lambda_{\mathbf{S}} \in \mathcal{R}} \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon)$ to be feasible, its optimal value would be attained for some $0 \leq \lambda_{\mathbf{S}}^* \in \mathcal{R}$, with $\lambda_{\mathbf{S}}^2 > \lambda_{\mathbf{S}}^*$. In this case, since $\lambda_{\mathbf{S}}^2 \in \mathcal{C}$ and by assumption $\lambda_{\mathbf{S}}^2 \geq 0$, it readily follows that $\lambda_{\mathbf{S}}^2 < \sqrt{\epsilon^2 - 1} / \epsilon$. On the other hand, since $0 \leq \lambda_{\mathbf{S}}^* \in \mathcal{R}$ by definition of the feasible set of the respective optimization problem, it must be true that $\lambda_{\mathbf{S}}^* \geq \sqrt{\epsilon^2 - 1} / \epsilon$. This, however, contradicts the fact that, also by definition, $\lambda_{\mathbf{S}}^2 > \lambda_{\mathbf{S}}^*$. Therefore, the problem $\max_{0 \leq \lambda_{\mathbf{S}} \in \mathcal{R}} \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon)$ is infeasible, returning by convention the value $-\infty$. In the case where $\lambda_{\mathbf{S}}^2 \in \mathcal{R}$, it is true that $\max_{0 \leq \lambda_{\mathbf{S}} \in \mathcal{R}} \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) \equiv \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^2, \epsilon)$. As above, if $\max_{\lambda_{\mathbf{S}} \in \mathcal{C}} \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) \right|$ is feasible, its optimal value $\left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^*, \epsilon) \right|$ would be attained for some $\lambda_{\mathbf{S}}^* \in \mathcal{C}$, with $\lambda_{\mathbf{S}}^2 > \lambda_{\mathbf{S}}^*$. Via a simple equivalence test, it can then be easily shown that $\lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^2, \epsilon) > \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^*, \epsilon) \right|$. Putting it altogether, it follows that the inner optimization problem in (75) equals $\left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^2, \epsilon) \right|$, for all $\epsilon > 0$. Therefore, and taking into account the facts about comparing $\left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}, \epsilon) \right|$ with $\epsilon / (1 + \epsilon)$, for $\lambda_{\mathbf{S}} \geq 0$ (since we have assumed that $\lambda_{\mathbf{S}}^2 \geq 0$), discussed above, we get that $\rho \left(\epsilon, \lambda_{\mathbf{S}}^2, \dots, \lambda_{\mathbf{S}}^S \right) \equiv \rho \left(\epsilon, \lambda_{\mathbf{S}}^2 \right)$, with

$$\rho \left(\epsilon, \lambda_{\mathbf{S}}^2 \right) \equiv \max \left\{ \frac{\epsilon}{1 + \epsilon}, \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^2, \epsilon) \right| \right\}$$

$$= \begin{cases} \left| \lambda_M^+(\lambda_S^2, \epsilon) \right|, & \text{if } \lambda_S^2 \in \mathcal{R} \vee \left[\lambda_S^2 \in \mathcal{C}, \epsilon \leq \frac{1 + \lambda_S^2}{1 - \lambda_S^2} \right], \\ \frac{\epsilon}{1 + \epsilon}, & \text{if } \lambda_S^2 \in \mathcal{C}, \epsilon > \frac{1 + \lambda_S^2}{1 - \lambda_S^2} \end{cases}, \quad (76)$$

which of course coincides with (18).

Suppose that $\lambda_S^2 < 0$. In this unlikely event, except for λ_S^1 , all other eigenvalues of \mathbf{S} will be negative. Instead of proceeding as above, we first observe that if $\epsilon \leq 1$, it will be true that $\lambda_S \in \mathcal{R}$, for any $\lambda_S \in [-1, 0)$. Second, under these circumstances, it can be easily shown that $\left| \lambda_M^+(\lambda_S, \epsilon) \right| \equiv \lambda_M^+(\lambda_S, \epsilon) > \left| \lambda_M^-(\lambda_S, \epsilon) \right|$ (due to the analysis presented above) and that $\lambda_M^+(\lambda_S, \epsilon)$ is strictly increasing when $\epsilon < 1$, whereas $\lambda_M^+(\lambda_S, \epsilon) \equiv \epsilon/(\epsilon + 1) \equiv 1/2$, if and only if $\epsilon \equiv 1$. Then, again via a careful equivalence test, it may be shown that $\left| \lambda_M^+(\lambda_S, \epsilon) \right| \equiv \lambda_M^+(\lambda_S, \epsilon) \geq \epsilon/(\epsilon + 1)$ if $\epsilon \leq 1$, and the complementary if $\epsilon > 1$. Thus, what remains is to study how $\left| \lambda_M^\pm(\lambda_S, \epsilon) \right|$ compare with $\epsilon/(\epsilon + 1)$ when $\lambda_S \in \mathcal{C}$ and $\epsilon > 1$. From the generic equivalence (73), it follows that, simply, $\left| \lambda_M^\pm(\lambda_S, \epsilon) \right| < \epsilon/(\epsilon + 1)$, whenever $\lambda_S \in (-1, 0)$. Also, when $\lambda_S \equiv -1$, $\left| \lambda_M^\pm(\lambda_S, \epsilon) \right| \leq 1/(\epsilon + 1)$, which is strictly smaller than $\epsilon/(\epsilon + 1)$, since $\epsilon > 1$. Combining all the above with the strict monotonicity of $\left| \lambda_M^\pm(\lambda_S, \epsilon) \right|$ when $\lambda_S \in \mathcal{C}$ and working similarly to (75), we again have $\rho(\epsilon, \lambda_S^2, \dots, \lambda_S^S) \equiv \rho(\epsilon, \lambda_S^2)$, with

$$\begin{aligned} \rho(\epsilon, \lambda_S^2) &\equiv \max \left\{ \frac{\epsilon}{1 + \epsilon}, \left| \lambda_M^+(\lambda_S^2, \epsilon) \right| \right\} \\ &= \begin{cases} \left| \lambda_M^+(\lambda_S^2, \epsilon) \right|, & \text{if } \epsilon \leq 1 \text{ (and } \lambda_S \in \mathcal{R}) \\ \frac{\epsilon}{1 + \epsilon}, & \text{if } \epsilon > 1 \end{cases}, \end{aligned} \quad (77)$$

which coincides with the respective expression of Theorem 2.

We now turn our attention to the problem of minimizing the SLEM of \mathbf{M} with respect to ϵ . Given what was stated above, we are interested in solving the minimax problem

$$\rho^*(\lambda_S^2) \triangleq \min_{\epsilon > 0} \rho(\epsilon, \lambda_S^2) \equiv \min_{\epsilon > 0} \max \left\{ \frac{\epsilon}{1 + \epsilon}, \left| \lambda_M^+(\lambda_S^2, \epsilon) \right| \right\}, \quad (78)$$

for given $\lambda_S^2 \in (-1, 1)$. It is of course reasonable to consider the cases where $\lambda_S^2 \geq 0$ and $\lambda_S^2 < 0$ separately, as above.

Suppose that $\lambda_S^2 \geq 0$. If additionally $\lambda_S^2 \in \mathcal{R}$ (see (76)), $\rho(\epsilon, \lambda_S) \equiv \lambda_M^+(\lambda_S^2, \epsilon)$. In the following, we show that $\rho(\epsilon, \lambda_S^2)$ is strictly decreasing in ϵ , for all $\lambda_S^2 \in \mathcal{R}$, through an equivalence test (actually, *either λ_S^2 is positive or not*). Suppose that $\partial \rho(\epsilon, \lambda_S^2) / \partial \epsilon < 0$. Then, after some algebra, this statement can be shown to be equivalent to

$$\lambda_S^2 \sqrt{1 + \epsilon^2 \left((\lambda_S^2)^2 - 1 \right)} < 1 - \epsilon \left((\lambda_S^2)^2 - 1 \right). \quad (79)$$

Since both sides of the inequality above are nonnegative for $\lambda_{\mathbf{S}}^2 \geq 0$ (if $\lambda_{\mathbf{S}}^2 < 0$, then take everything on the LHS and we are done), we can take squares and setting $x \triangleq \left(\lambda_{\mathbf{S}}^2\right)^2 - 1$, we get the equivalent statement $x(1+\epsilon)^2 < 0$, which is always true, since $x < 0$, for all $\lambda_{\mathbf{S}}^2 \in [0, 1)$. Thus, $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$ is strictly decreasing in ϵ . Through similar procedures, we can easily show that if $\lambda_{\mathbf{S}}^2 \in \mathcal{C}$, then $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$ is strictly increasing (for both respective subcases). Consequently, for $\lambda_{\mathbf{S}}^2 \geq 0$, $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$ is minimized at $\epsilon^*(\lambda_{\mathbf{S}}^2) \equiv 1/\sqrt{1 - \left(\lambda_{\mathbf{S}}^2\right)^2}$, with optimal value $\rho^*(\lambda_{\mathbf{S}}^2) \equiv \rho(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$.

Finally, suppose that $\lambda_{\mathbf{S}}^2 < 0$. From the previous discussion and (77), it follows that $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$ is strictly decreasing if $\epsilon \leq 1$ and strictly increasing otherwise. Therefore, for $\lambda_{\mathbf{S}}^2 < 0$, $\epsilon^*(\lambda_{\mathbf{S}}^2) \equiv 1$, with $\rho^*(\lambda_{\mathbf{S}}^2) \equiv 1/2$. The proof is complete. \blacksquare

Appendix B: Proof of Theorem 3

From Theorems 1 and 2, it is apparent that, in order to study the behavior of the consensus error bound with respect to the regularizer ϵ and the doubly stochastic matrix \mathbf{S} , it suffices to look at the bivariate function

$$f_n(\epsilon, \lambda_{\mathbf{S}}^2) \triangleq \frac{\left(\rho(\epsilon, \lambda_{\mathbf{S}}^2)\right)^{n-1}}{1+\epsilon}, \quad \epsilon > 0, \lambda_{\mathbf{S}}^2 \in [-1, 1), \quad (80)$$

parametrized by $n \in \mathbb{N}^2 \cap \mathbb{N}^{[\epsilon+1]}$. Whenever it exists, the first order derivative of $f_n(\epsilon, \lambda_{\mathbf{S}}^2)$ with respect to ϵ is given by

$$\frac{\partial f_n(\epsilon, \lambda_{\mathbf{S}}^2)}{\partial \epsilon} \equiv f_n(\epsilon, \lambda_{\mathbf{S}}^2) \left(\frac{(n-1)}{\rho(\epsilon, \lambda_{\mathbf{S}}^2)} \frac{\partial \rho(\epsilon, \lambda_{\mathbf{S}}^2)}{\partial \epsilon} - \frac{1}{(1+\epsilon)} \right). \quad (81)$$

From Theorem 2, as well as its proof presented in Appendix A above, we know that (recall the definitions of sets \mathcal{R} and \mathcal{C}) either (a) $\lambda_{\mathbf{S}}^2 \in \mathcal{R}$ and $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$ is strictly decreasing in ϵ , or (b) $\lambda_{\mathbf{S}}^2 \in \mathcal{C}$ and $\rho(\epsilon, \lambda_{\mathbf{S}}^2) \equiv \left| \lambda_{\mathbf{M}}^+(\lambda_{\mathbf{S}}^2, \epsilon) \right|$ is strictly increasing in ϵ , or (c) $\lambda_{\mathbf{S}}^2 \in \mathcal{C}$ and $\rho(\epsilon, \lambda_{\mathbf{S}}^2) \equiv \epsilon/(1+\epsilon)$, also strictly increasing in ϵ . For (a), things are trivial, since $\partial \rho(\epsilon, \lambda_{\mathbf{S}}^2)/\partial \epsilon < 0$ and, thus, $\partial f_n(\epsilon, \lambda_{\mathbf{S}}^2)/\partial \epsilon < 0$ as well, implying that $f_n(\epsilon, \lambda_{\mathbf{S}}^2)$ is strictly decreasing in ϵ , for all $n \in \mathbb{N}^2 \cap \mathbb{N}^{[\epsilon+1]}$. Now, for both (b) and (c), since $\partial \rho(\epsilon, \lambda_{\mathbf{S}}^2)/\partial \epsilon > 0$, (81) implies that

$$\frac{\partial f_n(\epsilon, \lambda_{\mathbf{S}}^2)}{\partial \epsilon} > 0 \Leftrightarrow n > \frac{\rho(\epsilon, \lambda_{\mathbf{S}}^2)}{(1+\epsilon) \partial \rho(\epsilon, \lambda_{\mathbf{S}}^2)/\partial \epsilon} + 1, \quad (82)$$

from where it can be carefully shown that the inequality on the right becomes $n > 2\epsilon + 1$ for case (b) and $n > \epsilon + 1$ for case (c). Therefore, choosing $n > 2\epsilon + 1$, for any ϵ results in a

convenient, global constraint for n . Next, fix $\mathbf{S} \in \mathfrak{S}$ and choose an $\epsilon_{max} \geq \epsilon^*(\lambda_{\mathbf{S}}^2)$. Then, as long as $n > 2\epsilon_{max} + 1$, it may be readily shown that $f_n(\epsilon, \lambda_{\mathbf{S}}^2)$ behaves exactly like $\rho(\epsilon, \lambda_{\mathbf{S}}^2)$ whenever $\epsilon \in (0, (n-1)/2]$, as far as where its extrema are located, as well as their type. In particular, since $\epsilon^*(\lambda_{\mathbf{S}}^2) \leq \epsilon_{max} < (n-1)/2$ by construction, it is true that $f_n(\epsilon, \lambda_{\mathbf{S}}^2)$ is minimized exactly at $\epsilon^*(\lambda_{\mathbf{S}}^2)$ (as defined in Theorem 2), whenever $\epsilon \in (0, \epsilon_{max}]$, *regardless of n* . Observe that demanding $n > 2\epsilon_{max} + 1$ implies that the constraint $n \in \mathbb{N}^2 \cap \mathbb{N}^{[\epsilon+1]}$ is satisfied, for all $\epsilon \in (0, \epsilon_{max}]$.

The case where $\epsilon > \epsilon_{max}$ is quite more complicated. To this end, let us consider the *multiobjective*, scalar, constrained optimization problem (recall the constraint $n \in \mathbb{N}^2 \cap \mathbb{N}^{[\epsilon+1]}$ and that $n > 2\epsilon_{max} + 1 > 2$)

$$\underset{\epsilon > 0}{\text{minimize}} \quad \left[\left\{ f_n(\epsilon, \lambda_{\mathbf{S}}^2) \text{ s.t. } \epsilon < n \right\}_{n > 2\epsilon_{max} + 1} \right]. \quad (83)$$

In order for the problem to be feasible, every individual objective must be feasible. Therefore, in order to satisfy the constraint $\epsilon < n$, *simultaneously for all $n > 2\epsilon_{max} + 1$* , it must be true that $\epsilon \in (0, \lfloor 2\epsilon_{max} + 2 \rfloor)$. Now, since the case $\epsilon \in (0, \epsilon_{max}]$ has already been covered above, suppose that $\epsilon \in (\epsilon_{max}, \lfloor 2\epsilon_{max} + 2 \rfloor)$. For such ϵ 's, let us compare the values of the objectives of (83) with those obtained for $\epsilon \in (0, \epsilon_{max}]$, and, in particular, for $\epsilon \equiv \epsilon^*(\lambda_{\mathbf{S}}^2)$. For any $n > 2\epsilon_{max} + 1$ and whenever $(n-1)/2 < \lfloor 2\epsilon_{max} + 2 \rfloor$, there are two possible choices for ϵ : either $\epsilon \in (\epsilon_{max}, (n-1)/2]$, where, due to monotonicity, $f_n(\epsilon, \lambda_{\mathbf{S}}^2) > f_n(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$, or $\epsilon \in ((n-1)/2, \lfloor 2\epsilon_{max} + 2 \rfloor)$, which might indeed be such that $f_n(\epsilon, \lambda_{\mathbf{S}}^2) < f_n(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$, since, in this region and for the particular n , $f_n(\epsilon, \lambda_{\mathbf{S}}^2)$ might be decreasing. However, for the same ϵ , there always exists $\tilde{n} > 2\epsilon_{max} + 1$ sufficiently large, corresponding to another objective of (83), such that $\epsilon \in (\epsilon_{max}, (\tilde{n}-1)/2]$, implying, as above, that $f_{\tilde{n}}(\epsilon, \lambda_{\mathbf{S}}^2) > f_{\tilde{n}}(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$. In the remaining case where $(n-1)/2 \geq \lfloor 2\epsilon_{max} + 2 \rfloor$, again from monotonicity as above, it readily follows that $f_n(\epsilon, \lambda_{\mathbf{S}}^2) > f_n(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$. We have shown that if, for some $n > 2\epsilon_{max} + 1$, there exists $\epsilon > 0$ such that $f_n(\epsilon, \lambda_{\mathbf{S}}^2) < f_n(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$, then there exists another $\tilde{n} > 2\epsilon_{max} + 1$, such that, for the same ϵ , the corresponding objective is relatively increased, that is, $f_{\tilde{n}}(\epsilon, \lambda_{\mathbf{S}}^2) > f_{\tilde{n}}(\epsilon^*(\lambda_{\mathbf{S}}^2), \lambda_{\mathbf{S}}^2)$. Therefore, the point $\epsilon \equiv \epsilon^*(\lambda_{\mathbf{S}}^2)$ constitutes a Pareto optimal solution for (83) and the resulting optimal bound (24) of Theorem 3 readily follows.

Regarding the second part of Theorem 3, for a fixed, “worst case” $\tilde{\mathbf{S}} \in \mathfrak{S}$, we focus on the function(al) (see (24))

$$g_n(\lambda_{\mathbf{S}}^2) \triangleq \gamma(\lambda_{\mathbf{S}}^2) \left(\rho^*(\lambda_{\mathbf{S}}^2) \right)^n, \quad \lambda_{\mathbf{S}}^2 \in [-1, 1), \quad (84)$$

for $n > 2\epsilon_{max} + 1 \geq 2\epsilon^*(\lambda_{\mathbf{S}}^2) + 1$. Via a simple but tedious equivalence test, it can be shown that

$$\frac{\partial g_n(\lambda_{\mathbf{S}}^2)}{\partial \lambda_{\mathbf{S}}^2} > 0 \Leftrightarrow n > \frac{\lambda_{\mathbf{S}}^2 + 1}{\sqrt{1 - (\lambda_{\mathbf{S}}^2)^2}}, \quad (85)$$

for all $\lambda_{\mathbf{S}}^2 \in [0, 1)$ (note that for $\lambda_{\mathbf{S}}^2 \in [-1, 0)$, $g_n(\lambda_{\mathbf{S}}^2)$ is constant), where it can be easily verified that the RHS of the above stated inequality on n (on the right) is itself strictly increasing in $\lambda_{\mathbf{S}}^2$ and that, additionally (recall that, by definition, $\epsilon_{max} \geq \epsilon^*(\lambda_{\mathbf{S}}^2)$), the already existent requirement that $n > 2\epsilon_{max} + 1$ implies the aforementioned inequality, regardless of $\lambda_{\mathbf{S}}^2 \in [0, \lambda_{\mathbf{S}}^2]$. This may be shown, again, via a, trivial this time, equivalence test. Then, it is guaranteed that $g_n(\lambda_{\mathbf{S}}^2)$ will be increasing whenever $\mathbf{S} \in \mathfrak{S}$ is additionally such that $\lambda_{\mathbf{S}}^2 \in [-1, \lambda_{\mathbf{S}}^2]$. Our claims follow. \blacksquare

Appendix C: Proof of Lemma 3

Simply, since

$$\left\| \boldsymbol{\vartheta}_t^j(n) - \boldsymbol{\vartheta}_t^j(\infty) \right\|_2^2 \equiv \sum_{k \in \mathbb{N}_S^+} \left| \boldsymbol{\vartheta}_t^j(n, k) - \theta_t(\mathbf{x}_j) \right|^2, \quad (86)$$

under the respective hypotheses, Lemma 1 implies that

$$\left| \boldsymbol{\vartheta}_t^j(n, k) - \theta_t(\mathbf{x}_j) \right|^2 \leq (\gamma \|\boldsymbol{\theta}_t(\mathbf{x}_j)\|_2 n \rho^n)^2, \quad (87)$$

for all $k \in \mathbb{N}_S^+$, which proves the result. \blacksquare

Appendix D: Proof of Lemma 4

By definition, it is true that

$$\begin{aligned} \frac{\theta_t^k(\mathbf{x}_j)}{S} &\equiv \left(\bar{\mathbf{y}}_t^k(\mathbf{x}_j) \right)^T \left(\boldsymbol{\Sigma}_t^k(\mathbf{x}_j) \right)^{-1} \bar{\mathbf{y}}_t^k(\mathbf{x}_j) + \log \det \left(\boldsymbol{\Sigma}_t^k(\mathbf{x}_j) \right) \\ &\leq \left\| \mathbf{y}_t^k - \boldsymbol{\mu}_t^k(\mathbf{x}_j) \right\|_2^2 \lambda_{inf} + N_k \log(\lambda_{sup}) \\ &\leq \left(\left\| \mathbf{y}_t^k \right\|_2 + \mu_{sup} \right)^2 \lambda_{inf} + N_k \log(\lambda_{sup}) \\ &\leq (\|\mathbf{y}_t\|_2 + \mu_{sup})^2 \lambda_{inf} + N \log(\lambda_{sup}). \end{aligned} \quad (88)$$

However, from Lemma 2, $\|\mathbf{y}_t\|_2 < \sqrt{\beta CN(1 + \log(T+1))}$ simultaneously for all $t \in \mathbb{N}_T$, with probability at least $1 - (T+1)^{1-CN} \exp(-CN)$. Consequently, in that event,

$$\frac{\theta_t^k(\mathbf{x}_j)}{S} \leq \left(\sqrt{\beta CN(1 + \log(T+1))} + \mu_{sup} \right)^2 \lambda_{inf} + N \log(\lambda_{sup}) \quad (89)$$

and since the bound $\sqrt{\beta CN(1 + \log(t+1))}$ is greater than unity, it will be true that

$$\frac{\theta_t^k(\mathbf{x}_j)}{S} \leq \beta CN(1 + \log(T+1)) \left((1 + \mu_{sup})^2 \lambda_{inf} + \log(\lambda_{sup}) \right). \quad (90)$$

Further, defining $\delta \triangleq \beta \left((1 + \mu_{sup})^2 \lambda_{inf} + \log(\lambda_{sup}) \right) > 1$, the ℓ_2 -norm of $\boldsymbol{\theta}_t(\mathbf{x}_j)$ can be upper bounded as

$$\begin{aligned} \|\boldsymbol{\theta}_t(\mathbf{x}_j)\|_2 &\equiv \sqrt{\sum_{k \in \mathcal{C}} \left(\theta_t^k(\mathbf{x}_j) \right)^2} \\ &\leq \delta S^{1.5} C N (1 + \log(T + 1)), \end{aligned} \quad (91)$$

with probability at least $1 - (T + 1)^{1 - CN} \exp(-CN)$. \blacksquare

Appendix E: Proof of Lemma 5

From Lemma 3, we know that

$$\boldsymbol{\vartheta}_t^j(n, k) \geq \theta_t(\mathbf{x}_L^j) - \gamma \|\boldsymbol{\theta}_t(\mathbf{x}_j)\|_2 n \rho^n, \quad (92)$$

from where, invoking Lemma 4 and by definition of $\theta_t(\mathbf{x}_j)$, we arrive at the lower bound

$$\boldsymbol{\vartheta}_t^j(n, k) \geq N \log(\lambda_{inf}) - \gamma \delta S^{1.5} C N (1 + \log(T + 1)) n \rho^n, \quad (93)$$

being true with measure at least $1 - (T + 1)^{1 - CN} \exp(-CN)$, under the respective hypotheses. Now, in order to bound $\boldsymbol{\vartheta}_t^j(n, k)$ from below, it suffices to demand that

$$N \log(\lambda_{inf}) - \gamma \delta S^{1.5} C N (1 + \log(T + 1)) n \rho^n \geq \frac{N}{2} \log(\lambda_{inf}). \quad (94)$$

Rearranging terms and taking logarithms, the above inequality will be true with probability at least $1 - (T + 1)^{1 - CN} \exp(-CN)$, provided that (36) holds, which constitutes needed to be shown. \blacksquare

References

- [1] D. S. Kalogieras and A. P. Petropulu, "Distributed nonlinear filtering of partially observed markov chains over WSNs: Truncating the ADMM," in *49th Asilomar Conference on Signals, Systems and Computers*, (Asilomar Hotel & Conference Grounds, Pacific Grove, CA, USA), November 2015.
- [2] O. Hlinka, F. Hlawatsch, and P. Djuric, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *Signal Processing Magazine, IEEE*, vol. 30, pp. 61–81, Jan 2013.
- [3] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, and D. Moore, "Environmental wireless sensor networks," *Proceedings of the IEEE*, vol. 98, pp. 1903–1917, Nov 2010.
- [4] J. Ko, C. Lu, M. Srivastava, J. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, vol. 98, pp. 1947–1960, Nov 2010.
- [5] T. Zhao and A. Nehorai, "Distributed sequential bayesian estimation of a diffusive source in wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 55, pp. 1511–1524, April 2007.

- [6] H. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*. Academic press, 2009.
- [7] F. Zhao and L. J. Guibas, *Wireless sensor networks: an information processing approach*. Morgan Kaufmann, 2004.
- [8] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless sensor networks for habitat monitoring,” in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, WSNA ’02, (New York, NY, USA), pp. 88–97, ACM, 2002.
- [9] R. Olfati-Saber and R. Murray, “Consensus problems in networks of agents with switching topology and time-delays,” *Automatic Control, IEEE Transactions on*, vol. 49, pp. 1520–1533, Sept 2004.
- [10] R. Olfati-Saber and J. Shamma, “Consensus filters for sensor networks and distributed sensor fusion,” in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC ’05. 44th IEEE Conference on*, pp. 6698–6703, Dec 2005.
- [11] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *Information Theory, IEEE Transactions on*, vol. 52, pp. 2508–2530, June 2006.
- [12] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proceedings of the IEEE*, vol. 98, pp. 1847–1864, Nov 2010.
- [13] B. Oreshkin, M. Coates, and M. Rabbat, “Optimization and analysis of distributed averaging with short node memory,” *Signal Processing, IEEE Transactions on*, vol. 58, pp. 2850–2865, May 2010.
- [14] T. Erseghe, D. Zennaro, E. Dall’Anese, and L. Vangelista, “Fast consensus by the alternating direction multipliers method,” *Signal Processing, IEEE Transactions on*, vol. 59, pp. 5523–5537, Nov 2011.
- [15] R. Olfati-Saber, “Distributed kalman filter with embedded consensus filters,” in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC ’05. 44th IEEE Conference on*, pp. 8179–8184, Dec 2005.
- [16] D. Spanos, R. Olfati-Saber, and R. Murray, “Approximate distributed kalman filtering in sensor networks with quantifiable performance,” in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pp. 133–139, April 2005.
- [17] R. Olfati-Saber, “Distributed kalman filtering for sensor networks,” in *Decision and Control, 2007 46th IEEE Conference on*, pp. 5492–5498, Dec 2007.
- [18] A. Ribeiro, G. Giannakis, and S. Roumeliotis, “Soi-kf: Distributed kalman filtering with low-cost communications using the sign of innovations,” *Signal Processing, IEEE Transactions on*, vol. 54, pp. 4782–4795, Dec 2006.
- [19] I. Schizas, G. Giannakis, S. Roumeliotis, and A. Ribeiro, “Consensus in ad hoc wsns with noisy links—part ii: Distributed estimation and smoothing of random signals,” *Signal Processing, IEEE Transactions on*, vol. 56, pp. 1650–1666, April 2008.

- [20] P. Alriksson and A. Rantzer, “Distributed kalman filtering using weighted averaging,” in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, pp. 2445–2450, 2006.
- [21] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, “Distributed kalman filtering based on consensus strategies,” *Selected Areas in Communications, IEEE Journal on*, vol. 26, pp. 622–633, May 2008.
- [22] M. Mahmoud and H. Khalid, “Distributed kalman filtering: a bibliographic review,” *Control Theory Applications, IET*, vol. 7, pp. 483–501, March 2013.
- [23] S. Bandyopadhyay and S.-J. Chung, “Distributed estimation using bayesian consensus filtering,” in *American Control Conference (ACC), 2014*, pp. 634–641, June 2014.
- [24] R. J. Elliott, “Exact adaptive filters for markov chains observed in gaussian noise,” *Automatica*, vol. 30, no. 9, pp. 1399–1408, 1994.
- [25] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models*. Springer, 1994.
- [26] J. Dall and M. Christensen, “Random geometric graphs,” *Physical Review E*, vol. 66, no. 1, p. 016121, 2002.
- [27] M. Penrose, *Random geometric graphs*, vol. 5. Oxford University Press Oxford, 2003.
- [28] N. Ghasemi, S. Dey, and J. Baras, “Stochastic average consensus filter for distributed HMM filtering,” tech. rep., Institute for Systems Research (ISR), May 2010.
- [29] S. Boyd, P. Diaconis, and L. Xiao, “Fastest mixing markov chain on a graph,” *SIAM review*, vol. 46, no. 4, pp. 667–689, 2004.
- [30] D. S. Kalogerias and A. P. Petropulu, “Asymptotically optimal discrete time nonlinear filters from stochastically convergent state process approximations,” *IEEE Transactions on Signal Processing*, vol. 63, pp. 3522 – 3536, July 2015.
- [31] D. S. Kalogerias and A. P. Petropulu, “Grid-based filtering of Markov processes revisited: Recursive estimation & asymptotic optimality,” *Draft available at: <http://eceweb1.rutgers.edu/~dkalogerias/>*, 2015.
- [32] J. R. Silvester, “Determinants of block matrices,” *The Mathematical Gazette*, vol. 84, no. 501, pp. 460–467, 2000.